Supplementary information for "Enforcing stationarity through the prior in vector autoregressions"

Sarah E. Heaps
Department of Mathematical Sciences, Durham University

May 13, 2022

Abstract

Supplementary material contained in this note includes proofs of the forward and reverse mappings between the original parameters of the VAR model and the new set of uncontrained parameters; additional plots, simulations and text illustrating properties of the prior over the stationary region; full details of the modified parameterization of Roy et al. (2019) and its vague prior; a complete description of the extension of the reparameterization and prior to VARMA models; and further details on the application to macroeconomic data.

S1 Proofs of mappings between the $\mathrm{VAR}_m(p)$ parameters and partial autocorrelations

S1.1 Definitions and preliminary results

Let G_s and g_s $(s = 1, \ldots, p)$ be $ms \times ms$ and $ms \times m$ block matrices defined as

$$G_s = \begin{pmatrix} \varGamma_0 & \varGamma_1^{\mathrm{T}} & \cdots & \varGamma_{s-1}^{\mathrm{T}} \\ \varGamma_1 & \varGamma_0 & \cdots & \varGamma_{s-2}^{\mathrm{T}} \\ \vdots & \vdots & \ddots & \vdots \\ \varGamma_{s-1} & \varGamma_{s-2} & \cdots & \varGamma_0 \end{pmatrix}, \qquad g_s = \begin{pmatrix} \varGamma_1 \\ \varGamma_2 \\ \vdots \\ \varGamma_s \end{pmatrix}$$

with blockwise transposes denoted by

$$G_s^* = \begin{pmatrix} \varGamma_0 & \varGamma_1 & \cdots & \varGamma_{s-1} \\ \varGamma_1^{\mathrm{T}} & \varGamma_0 & \cdots & \varGamma_{s-2} \\ \vdots & \vdots & \ddots & \vdots \\ \varGamma_{s-1}^{\mathrm{T}} & \varGamma_{s-2}^{\mathrm{T}} & \cdots & \varGamma_0 \end{pmatrix}, \qquad g_s^* = \begin{pmatrix} \varGamma_1^{\mathrm{T}} \\ \varGamma_2^{\mathrm{T}} \\ \vdots \\ \varGamma_s^{\mathrm{T}} \end{pmatrix}.$$

Now denote by $\tilde{g}_s$ and $\tilde{g}_s^*$ the reversed matrices

$$\tilde{g}_s = Qg_s = \begin{pmatrix} \varGamma_s \\ \varGamma_{s-1} \\ \vdots \\ \varGamma_1 \end{pmatrix}, \qquad \tilde{g}_s^* = Qg_s^* = \begin{pmatrix} \varGamma_s^{\mathrm{T}} \\ \varGamma_{s-1}^{\mathrm{T}} \\ \vdots \\ \varGamma_1^{\mathrm{T}} \end{pmatrix}$$

in which the $ms \times ms$ block matrix Q, defined by

$$Q = \begin{pmatrix} 0_m & \cdots & 0_m & I_m \\ 0_m & \cdots & I_m & 0_m \\ \vdots & \ddots & \vdots & \vdots \\ I_m & \cdots & 0_m & 0_m \end{pmatrix}$$

is involutory (i.e. its own inverse), symmetric and hence orthogonal.

Let $\boldsymbol{y}_{i:j} = (\boldsymbol{y}_i^{\mathrm{T}}, \ldots, \boldsymbol{y}_j^{\mathrm{T}})^{\mathrm{T}}$ and

$$\Phi_s = \begin{pmatrix} \phi_{s1}^{\mathrm{T}} \\ \vdots \\ \phi_{ss}^{\mathrm{T}} \end{pmatrix}, \qquad \Phi_{s,-s} = \begin{pmatrix} \phi_{s1}^{\mathrm{T}} \\ \vdots \\ \phi_{s,s-1}^{\mathrm{T}} \end{pmatrix}.$$

Similarly, define

$$\Phi_s^* = \begin{pmatrix} \phi_{s1}^{*T} \\ \vdots \\ \phi_{ss}^{*T} \end{pmatrix}, \qquad \Phi_{s,-s}^* = \begin{pmatrix} \phi_{s1}^{*T} \\ \vdots \\ \phi_{s,s-1}^{*T} \end{pmatrix}.$$

Using standard multivariate normal theory, the conditional mean of $\boldsymbol{y}_{t+1}$ given its s predecessors $\boldsymbol{y}_{t:t-s+1}$ is

$$\begin{aligned} E(\boldsymbol{y}_{t+1} \mid \boldsymbol{y}_{t:t-s+1}) &= E(\boldsymbol{y}_{t+1}) + \mathrm{Cov}(\boldsymbol{y}_{t+1}, \boldsymbol{y}_{t:t-s+1})\mathrm{Var}(\boldsymbol{y}_{t:t-s+1})^{-1}\left\{\boldsymbol{y}_{t:t-s+1} - E(\boldsymbol{y}_{t:t-s+1})\right\} \\ &= \mathbf{0} + g_s^{\mathrm{T}} G_s^{-1}(\boldsymbol{y}_{t:t-s+1} - \mathbf{0}) \\ &= g_s^{\mathrm{T}} G_s^{-1} \boldsymbol{y}_{t:t-s+1} \\ &= \Phi_s^{\mathrm{T}} \boldsymbol{y}_{t:t-s+1} \quad (s = 1, \ldots, p). \end{aligned}$$

Therefore

$$\Phi_s = G_s^{-1} g_s \quad \Longleftrightarrow \quad g_s = G_s \Phi_s, \quad (s = 1, \ldots, p), \tag{S1}$$

in which we refer to the equations on the right as the forward prediction equations. The corresponding conditional variance is

$$\begin{aligned} \mathrm{Var}(\boldsymbol{y}_{t+1} \mid \boldsymbol{y}_{t:t-s+1}) &= \mathrm{Var}(\boldsymbol{y}_{t+1}) - \mathrm{Cov}(\boldsymbol{y}_{t+1}, \boldsymbol{y}_{t:t-s+1})\mathrm{Var}(\boldsymbol{y}_{t:t-s+1})^{-1}\mathrm{Cov}(\boldsymbol{y}_{t:t-s+1}, \boldsymbol{y}_{t+1}) \\ &= \Gamma_0 - g_s^{\mathrm{T}} G_s^{-1} g_s \\ &= \Gamma_0 - g_s^{\mathrm{T}} \Phi_s \\ &= \Sigma_s \quad (s = 1, \ldots, p). \end{aligned}$$

By symmetry, this gives

$$\Sigma_s = \Gamma_0 - \Phi_s^{\mathrm{T}} g_s, \quad (s = 1, \ldots, p). \tag{S2}$$

Similarly, the conditional mean of $\boldsymbol{y}_{t-s}$ given its s successors $\boldsymbol{y}_{(t-s+1):t}$ is

$$\begin{aligned}
&E(\boldsymbol{y}_{t-s} \mid \boldsymbol{y}_{(t-s+1):t}) \\
&\quad = E(\boldsymbol{y}_{t-s}) + \mathrm{Cov}(\boldsymbol{y}_{t-s}, \boldsymbol{y}_{(t-s+1):t})\mathrm{Var}(\boldsymbol{y}_{(t-s+1):t})^{-1}\left\{\boldsymbol{y}_{(t-s+1):t} - E(\boldsymbol{y}_{(t-s+1):t})\right\} \\
&\quad = \mathbf{0} + g_s^{*T} G_s^{*-1}(\boldsymbol{y}_{(t-s+1):t} - \mathbf{0}) \\
&\quad = g_s^{*T} G_s^{*-1}\boldsymbol{y}_{(t-s+1):t} \\
&\quad = \Phi_s^{*T}\boldsymbol{y}_{(t-s+1):t} \quad (s = 1, \ldots, p).
\end{aligned}$$

Therefore

$$\Phi_s^* = G_s^{*-1} g_s^* \quad \Longleftrightarrow \quad g_s^* = G_s^* \Phi_s^*, \quad (s = 1, \ldots, p), \tag{S3}$$

in which we refer to the equations on the right as the reverse prediction equations. The corresponding conditional variance is

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{y}_{t-s} \mid \boldsymbol{y}_{(t-s+1):t}) &= \mathrm{Var}(\boldsymbol{y}_{t-s}) + \mathrm{Cov}(\boldsymbol{y}_{t-s}, \boldsymbol{y}_{(t-s+1):t})\mathrm{Var}(\boldsymbol{y}_{(t-s+1):t})^{-1}\mathrm{Cov}(\boldsymbol{y}_{(t-s+1):t}, \boldsymbol{y}_{t-s}) \\
&= \Gamma_0 - g_s^{*T} G_s^{*-1} g_s^* \\
&= \Gamma_0 - g_s^{*T}\Phi_s^* \\
&= \Sigma_s^* \quad (s = 1, \ldots, p).
\end{aligned}$$

By symmetry, this gives

$$\Sigma_s^* = \Gamma_0 - \Phi_s^{*T} g_s^*, \quad (s = 1, \ldots, p). \tag{S4}$$

Before proceeding with the proofs of the forward and reverse mapping, we use the properties of Q and the forward and reverse prediction equations to derive the following preliminary results

$$Q\Phi_s = QG_s^{-1} g_s = (QG_sQ)^{-1} Q g_s = G_s^{*-1}\tilde{g}_s \quad (s = 1, \ldots, p). \tag{S5}$$

Similarly,

$$Q\Phi_s^* = QG_s^{*-1} g_s^* = (QG_s^*Q)^{-1} Q g_s^* = G_s^{-1}\tilde{g}_s^* \quad (s = 1, \ldots, p). \tag{S6}$$

S1.2 Forward mapping

For $s = 1$, the forward prediction equation simplifies to give

$$g_1 = G_1 \Phi_1 \quad \Longleftrightarrow \quad \Gamma_1 = \Gamma_0 \phi_{11}^{\mathrm{T}}$$

and so

$$\phi_{11} = \Gamma_1^{\mathrm{T}} \Gamma_0^{-1} = \Gamma_1^{\mathrm{T}} \Sigma_0^{*\,-1}. \tag{S7}$$

Similarly, from the first reverse prediction equation, we arrive at

$$g_1^* = G_1^* \Phi_1^* \quad \Longleftrightarrow \quad \Gamma_1^{\mathrm{T}} = \Gamma_0 \phi_{11}^{*\,T}$$

and so

$$\phi_{11}^* = \Gamma_1 \Gamma_0^{-1} = \Gamma_1 \Sigma_0^{-1}. \tag{S8}$$

We can partition the matrices in the forward prediction equation, $G_s \Phi_s = g_s$ ($s = 2, \ldots, p$), as follows

$$\begin{pmatrix} G_{s-1} & \tilde{g}_{s-1}^* \\ \tilde{g}_{s-1}^{*\,T} & \Gamma_0 \end{pmatrix} \begin{pmatrix} \Phi_{s,-s} \\ \phi_{ss}^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} g_{s-1} \\ \Gamma_s \end{pmatrix}$$

or, equivalently,

$$G_{s-1} \Phi_{s,-s} + \tilde{g}_{s-1}^* \phi_{ss}^{\mathrm{T}} = g_{s-1}, \tag{S9}$$

$$\tilde{g}_{s-1}^{*\,T} \Phi_{s,-s} + \Gamma_0 \phi_{ss}^{\mathrm{T}} = \Gamma_s. \tag{S10}$$

Similarly, we can partition the matrices in the backward prediction equation, $G_s^* \Phi_s^* = g_s^*$ ($s = 2, \ldots, p$), as follows

$$\begin{pmatrix} G_{s-1}^* & \tilde{g}_{s-1} \\ \tilde{g}_{s-1}^{\mathrm{T}} & \Gamma_0 \end{pmatrix} \begin{pmatrix} \Phi_{s,-s}^* \\ \phi_{ss}^{*\,T} \end{pmatrix} = \begin{pmatrix} g_{s-1}^* \\ \Gamma_s^{\mathrm{T}} \end{pmatrix}$$

or, equivalently,

$$G_{s-1}^* \Phi_{s,-s}^* + \tilde{g}_{s-1} \phi_{ss}^{*\,T} = g_{s-1}^*, \tag{S11}$$

$$\tilde{g}_{s-1}^{\mathrm{T}} \Phi_{s,-s}^* + \Gamma_0 \phi_{ss}^{*\,T} = \Gamma_s^{\mathrm{T}}. \tag{S12}$$

From (S9) we can write

$$\Phi_{s,-s} = G_{s-1}^{-1}(g_{s-1} - \tilde{g}_{s-1}^{*}\phi_{ss}^{\mathrm{T}}) = G_{s-1}^{-1}g_{s-1} - G_{s-1}^{-1}\tilde{g}_{s-1}^{*}\phi_{ss}^{\mathrm{T}}$$

and then using (S1) and (S6) we have

$$\Phi_{s,-s} = \Phi_{s-1} - Q\Phi_{s-1}^{*}\phi_{ss}^{\mathrm{T}}. \tag{S13}$$

So, blockwise, we have

$$\phi_{s,i} = \phi_{s-1,i} - \phi_{ss}\phi_{s-1,s-i}^{*}, \quad (s = 2, \ldots, p;\ i = 1, \ldots, s-1). \tag{S14}$$

Similarly, from (S11) we can write

$$\Phi_{s,-s}^{*} = G_{s-1}^{*\,-1}(g_{s-1}^{*} - \tilde{g}_{s-1}\phi_{ss}^{*\,T}) = G_{s-1}^{*\,-1}g_{s-1}^{*} - G_{s-1}^{*\,-1}\tilde{g}_{s-1}\phi_{ss}^{*\,T}$$

and then using (S3) and (S5) we have

$$\Phi_{s,-s}^{*} = \Phi_{s-1}^{*} - Q\Phi_{s-1}\phi_{ss}^{*\,T}. \tag{S15}$$

So, blockwise, we have

$$\phi_{s,i}^{*} = \phi_{s-1,i}^{*} - \phi_{ss}^{*}\phi_{s-1,s-i}, \quad (s = 2, \ldots, p;\ i = 1, \ldots, s-1). \tag{S16}$$

Equations (S14) and (S16) establish the results of step 2(b)(ii) in the forward mapping from the Appendix in the paper.

Next, taking (S13) in (S10) and using properties of Q we can write

$$\begin{aligned}
\Gamma_0\phi_{ss}^{\mathrm{T}} &= \Gamma_s - \tilde{g}_{s-1}^{*\,T}\Phi_{s,-s} \\
&= \Gamma_s - \tilde{g}_{s-1}^{*\,T}(\Phi_{s-1} - Q\Phi_{s-1}^{*}\phi_{ss}^{\mathrm{T}}) \\
&= \Gamma_s - \tilde{g}_{s-1}^{*\,T}\Phi_{s-1} + (Qg_{s-1}^{*})^{\mathrm{T}}Q\Phi_{s-1}^{*}\phi_{ss}^{\mathrm{T}} \\
&= \Gamma_s - \tilde{g}_{s-1}^{*\,T}\Phi_{s-1} + g_{s-1}^{*\,T}\Phi_{s-1}^{*}\phi_{ss}^{\mathrm{T}}.
\end{aligned}$$

We therefore have

$$\phi_{ss}(\Gamma_0 - \Phi_{s-1}^{*\,T}g_{s-1}^{*}) = \Gamma_s^{\mathrm{T}} - \Phi_{s-1}^{\mathrm{T}}\tilde{g}_{s-1}^{*}$$

which, from (S4), is equal to

$$\phi_{ss}\Sigma^*_{s-1} = \Gamma_s^{\mathrm{T}} - \Phi_{s-1}^{\mathrm{T}}\tilde{g}^*_{s-1}. \tag{S17}$$

Solving for ϕ_{ss}, we can write

$$\phi_{ss} = (\Gamma_s^{\mathrm{T}} - \Phi_{s-1}^{\mathrm{T}}\tilde{g}^*_{s-1})\Sigma^{*-1}_{s-1} = (\Gamma_s^{\mathrm{T}} - \phi_{s-1,1}\Gamma_{s-1}^{\mathrm{T}} - \cdots - \phi_{s-1,s-1}\Gamma_1^{\mathrm{T}})\Sigma^{*-1}_{s-1}, \tag{S18}$$

for $s = 1, \ldots, p$ which encompasses (S7) as a special case in which $\Phi_0^{\mathrm{T}}\tilde{g}^*_0 \equiv 0$.

Similarly, taking (S15) in (S12) and using properties of Q we can write

$$\begin{aligned}
\Gamma_0\phi^{*T}_{ss} &= \Gamma_s^{\mathrm{T}} - \tilde{g}_{s-1}^{\mathrm{T}}\Phi^*_{s,-s} \\
&= \Gamma_s^{\mathrm{T}} - \tilde{g}_{s-1}^{\mathrm{T}}(\Phi^*_{s-1} - Q\Phi_{s-1}\phi^{*T}_{ss}) \\
&= \Gamma_s^{\mathrm{T}} - \tilde{g}_{s-1}^{\mathrm{T}}\Phi^*_{s-1} + (Qg_{s-1})^{\mathrm{T}}Q\Phi_{s-1}\phi^{*T}_{ss} \\
&= \Gamma_s^{\mathrm{T}} - \tilde{g}_{s-1}^{\mathrm{T}}\Phi^*_{s-1} + g_{s-1}^{\mathrm{T}}\Phi_{s-1}\phi^{*T}_{ss}.
\end{aligned}$$

We therefore have

$$\phi^*_{ss}(\Gamma_0 - \Phi_{s-1}^{\mathrm{T}}g_{s-1}) = \Gamma_s - \Phi^{*T}_{s-1}\tilde{g}_{s-1}$$

which, from (S2) is equal to

$$\phi^*_{ss}\Sigma_{s-1} = \Gamma_s - \Phi^{*T}_{s-1}\tilde{g}_{s-1}. \tag{S19}$$

Solving for ϕ^*_{ss}, we can write

$$\phi^*_{ss} = (\Gamma_s - \Phi^{*T}_{s-1}\tilde{g}_{s-1})\Sigma^{-1}_{s-1} = (\Gamma_s - \phi^*_{s-1,1}\Gamma_{s-1} - \cdots - \phi^*_{s-1,s-1}\Gamma_1)\Sigma^{-1}_{s-1}, \tag{S20}$$

for $s = 1, \ldots, p$ which encompasses (S8) as a special case in which $\Phi^{*T}_0\tilde{g}_0 \equiv 0$. Equations (S18) and (S20) establish the results in step 2(b)(i) of the forward mapping.

The recursion for the matrices of autoregressive coefficients in the sth forward and reverse prediction equations involve the conditional variances Σ_{s-1} and Σ^*_{s-1} ($s = 1, \ldots, p-1$) and so we need to define a recursion for their computation. Post-multiplying both sides of the left-hand equation in (1) by $\boldsymbol{y}_{t+1}^{\mathrm{T}}$ and taking expectations yields

$$E(\boldsymbol{y}_{t+1}\boldsymbol{y}_{t+1}^{\mathrm{T}}) = \sum_{i=1}^{s}\phi_{si}E(\boldsymbol{y}_{t-i+1}\boldsymbol{y}_{t+1}^{\mathrm{T}}) + E(\boldsymbol{\epsilon}_{s,t+1}\boldsymbol{y}_{t+1}^{\mathrm{T}}) \iff \Gamma_0 = \sum_{i=1}^{s}\phi_{si}\Gamma_i + \Sigma_s$$

and so

$$\Sigma_s = \Gamma_0 - \phi_{s1}\Gamma_1 - \cdots \phi_{ss}\Gamma_s, \quad (s = 1, \ldots, p). \tag{S21}$$

Similarly, post-multiplying both sides of the right-hand equation in (1) by $\boldsymbol{y}_{t-s}^{\mathrm{T}}$ and taking expectations yields

$$E(\boldsymbol{y}_{t-s}\boldsymbol{y}_{t-s}^{\mathrm{T}}) = \sum_{i=1}^{s} \phi_{si}^{*} E(\boldsymbol{y}_{t-s+i}\boldsymbol{y}_{t-s}^{\mathrm{T}}) + E(\boldsymbol{\epsilon}_{s,t-s}^{*}\boldsymbol{y}_{t-s}^{\mathrm{T}}) \iff \Gamma_0 = \sum_{i=1}^{s} \phi_{si}^{*}\Gamma_i^{\mathrm{T}} + \Sigma_s^{*}$$

and so

$$\Sigma_s^{*} = \Gamma_0 - \phi_{s1}^{*}\Gamma_1^{\mathrm{T}} - \cdots \phi_{ss}^{*}\Gamma_s^{\mathrm{T}}, \quad (s = 1, \ldots, p). \tag{S22}$$

Equations (S21) and (S22) establish the results of step 2(b)(iv) in the forward mapping.

Recall that we define the partial autocorrelation matrices $P_1, \ldots, P_p$ as $P_{s+1} = \mathrm{Cov}(\boldsymbol{z}_{s,t+1}, \boldsymbol{z}_{s,t-s}^{*})$ $(s = 0, \ldots, p-1)$. For $s = 0$, we can write this as

$$P_1 = \mathrm{Cov}(S_0^{-1}\boldsymbol{y}_{t+1}\,,\, S_0^{*\,-1}\boldsymbol{y}_t) = S_0^{-1}\mathrm{Cov}(\boldsymbol{y}_{t+1}, \boldsymbol{y}_t)(S_0^{*\,T})^{-1} = S_0^{-1}\Gamma_1^{\mathrm{T}}(S_0^{*\,T})^{-1}$$

which, using (S7), can be expressed as

$$P_1 = S_0^{-1}\phi_{11}\Sigma_0^{*}(S_0^{*\,T})^{-1} = S_0^{-1}\phi_{11}S_0^{*}S_0^{*\,T}(S_0^{*\,T})^{-1} = S_0^{-1}\phi_{11}S_0^{*}. \tag{S23}$$

Now, for $s = 1, \ldots, p-1$ we have

$$P_{s+1} = \mathrm{Cov}(S_s^{-1}\boldsymbol{\epsilon}_{s,t+1}\,,\, S_s^{*\,-1}\boldsymbol{\epsilon}_{s,t-s}^{*}) = S_s^{-1}\mathrm{Cov}(\boldsymbol{\epsilon}_{s,t+1}, \boldsymbol{\epsilon}_{s,t-s}^{*})(S_s^{*\,T})^{-1}.$$

We can write the inner covariance as

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{\epsilon}_{s,t+1}, \boldsymbol{\epsilon}_{s,t-s}^{*}) &= \mathrm{Cov}(\boldsymbol{y}_{t+1} - \Phi_s^{\mathrm{T}}\boldsymbol{y}_{t:t-s+1}\,,\, \boldsymbol{y}_{t-s} - \Phi_s^{*\,T}\boldsymbol{y}_{t-s+1:t}) \\
&= \mathrm{Cov}(\boldsymbol{y}_{t+1}, \boldsymbol{y}_{t-s}) - \mathrm{Cov}(\boldsymbol{y}_{t+1}, \boldsymbol{y}_{t-s+1:t})\Phi_s^{*} \\
&\quad - \Phi_s^{\mathrm{T}}\mathrm{Cov}(\boldsymbol{y}_{t:t-s+1}, \boldsymbol{y}_{t-s}) + \Phi_s^{\mathrm{T}}\mathrm{Cov}(\boldsymbol{y}_{t:t-s+1}, \boldsymbol{y}_{t-s+1:t})\Phi_s^{*} \\
&= \mathrm{Cov}(\boldsymbol{y}_{t+1}, \boldsymbol{y}_{t-s}) - \mathrm{Cov}(\boldsymbol{y}_{t+1}, Q\boldsymbol{y}_{t:t-s+1})\Phi_s^{*} \\
&\quad - \Phi_s^{\mathrm{T}}\mathrm{Cov}(\boldsymbol{y}_{t:t-s+1}, \boldsymbol{y}_{t-s}) + \Phi_s^{\mathrm{T}}\mathrm{Cov}(\boldsymbol{y}_{t:t-s+1}, Q\boldsymbol{y}_{t:t-s+1})\Phi_s^{*} \\
&= \Gamma_{s+1}^{\mathrm{T}} - g_s^{\mathrm{T}}Q\Phi_s^{*} - \Phi_s^{\mathrm{T}}\tilde{g}_s^{*} + \Phi_s^{\mathrm{T}}G_sQ\Phi_s^{*}.
\end{aligned}$$

Using (S6), this can be written as

$$\mathrm{Cov}(\boldsymbol{\epsilon}_{s,t+1}, \boldsymbol{\epsilon}_{s,t-s}^{*}) = \Gamma_{s+1}^{\mathrm{T}} - g_s^{\mathrm{T}}G_s^{-1}\tilde{g}_s^{*} - \Phi_s^{\mathrm{T}}\tilde{g}_s^{*} + \Phi_s^{\mathrm{T}}G_sG_s^{-1}\tilde{g}_s^{*}$$

which, using (S1), gives

$$\mathrm{Cov}(\boldsymbol{\epsilon}_{s,t+1}, \boldsymbol{\epsilon}^*_{s,t-s}) = \varGamma^{\mathrm{T}}_{s+1} - \varPhi^{\mathrm{T}}_s \tilde{g}^*_s.$$

Finally, from (S17) we obtain

$$\mathrm{Cov}(\boldsymbol{\epsilon}_{s,t+1}, \boldsymbol{\epsilon}^*_{s,t-s}) = \phi_{s+1,s+1}\varSigma^*_s.$$

It follows that

$$P_{s+1} = S_s^{-1}\phi_{s+1,s+1}\varSigma^*_s(S_s^{*T})^{-1} = S_s^{-1}\phi_{s+1,s+1}S^*_s S^{*T}_s(S^{*T}_s)^{-1}.$$

Therefore, also taking account of (S23), we have

$$P_{s+1} = S_s^{-1}\phi_{s+1,s+1}S^*_s, \quad (s = 0, \ldots, p-1). \tag{S24}$$

To obtain an alternative representation of the partial autocorrelation matrices, it is instructive to construct P^{T}_{s+1} $(s = 0, \ldots, p-1)$. For $s = 0$ we can write

$$P^{\mathrm{T}}_1 = \mathrm{Cov}(S^{*\,-1}_0 \boldsymbol{y}_t\,,\, S_0^{-1}\boldsymbol{y}_{t+1}) = S^{*\,-1}_0 \mathrm{Cov}(\boldsymbol{y}_t, \boldsymbol{y}_{t+1})(S_0^T)^{-1} = S^{*\,-1}_0 \varGamma_1 (S_0^T)^{-1}$$

which, using (S8), can be expressed as

$$P^{\mathrm{T}}_1 = S^{*\,-1}_0 \phi^*_{11}\varSigma_0(S_0^T)^{-1} = S^{*\,-1}_0 \phi^*_{11} S_0 S_0^{\mathrm{T}}(S_0^T)^{-1} = S^{*\,-1}_0 \phi^*_{11} S_0$$

and so

$$P_1 = (S^{*\,-1}_0 \phi^*_{11} S_0)^{\mathrm{T}}. \tag{S25}$$

Now, for $s = 1, \ldots, p-1$ we can write

$$P^{\mathrm{T}}_{s+1} = \mathrm{Cov}(S^{*\,-1}_s \boldsymbol{\epsilon}^*_{s,t-s}\,,\, S_s^{-1}\boldsymbol{\epsilon}_{s,t+1}) = S^{*\,-1}_s \mathrm{Cov}(\boldsymbol{\epsilon}^*_{s,t-s}, \boldsymbol{\epsilon}_{s,t+1})(S_s^{\mathrm{T}})^{-1}.$$

The inner covariance can be expressed as

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{\epsilon}^*_{s,t-s}, \boldsymbol{\epsilon}_{s,t+1}) &= \mathrm{Cov}(\boldsymbol{y}_{t-s} - \varPhi^{*T}_s \boldsymbol{y}_{t-s+1:t}\,,\, \boldsymbol{y}_{t+1} - \varPhi^{\mathrm{T}}_s \boldsymbol{y}_{t:t-s+1}) \\
&= \mathrm{Cov}(\boldsymbol{y}_{t-s}, \boldsymbol{y}_{t+1}) - \mathrm{Cov}(\boldsymbol{y}_{t-s}, \boldsymbol{y}_{t:t-s+1})\varPhi_s \\
&\quad - \varPhi^{*T}_s \mathrm{Cov}(\boldsymbol{y}_{t-s+1:t}, \boldsymbol{y}_{t+1}) + \varPhi^{*T}_s \mathrm{Cov}(\boldsymbol{y}_{t-s+1:t}, \boldsymbol{y}_{t:t-s+1})\varPhi_s \\
&= \mathrm{Cov}(\boldsymbol{y}_{t-s}, \boldsymbol{y}_{t+1}) - \mathrm{Cov}(\boldsymbol{y}_{t-s}, Q\boldsymbol{y}_{t-s+1:t})\varPhi_s \\
&\quad - \varPhi^{*T}_s \mathrm{Cov}(\boldsymbol{y}_{t-s+1:t}, \boldsymbol{y}_{t+1}) + \varPhi^{*T}_s \mathrm{Cov}(\boldsymbol{y}_{t-s+1:t}, Q\boldsymbol{y}_{t-s+1:t})\varPhi_s \\
&= \varGamma_{s+1} - g^{*T}_s Q\varPhi_s - \varPhi^{*T}_s \tilde{g}_s + \varPhi^{*T}_s G^*_s Q\varPhi_s.
\end{aligned}$$

Using (S5), this can be written as

$$\mathrm{Cov}(\boldsymbol{\epsilon}^*_{s,t-s}, \boldsymbol{\epsilon}_{s,t+1}) = \Gamma_{s+1} - g_s^{*T} G_s^{*-1} \tilde{g}_s - \Phi_s^{*T} \tilde{g}_s + \Phi_s^{*T} G_s^* G_s^{*-1} \tilde{g}_s$$

which, using (S3), gives

$$\mathrm{Cov}(\boldsymbol{\epsilon}^*_{s,t-s}, \boldsymbol{\epsilon}_{s,t+1}) = \Gamma_{s+1} - \Phi_s^{*T} \tilde{g}_s.$$

Finally, from (S19) we obtain

$$\mathrm{Cov}(\boldsymbol{\epsilon}^*_{s,t-s}, \boldsymbol{\epsilon}_{s,t+1}) = \phi^*_{s+1,s+1} \Sigma_s.$$

It follows that

$$P_{s+1}^{\mathrm{T}} = S_s^{*-1} \phi^*_{s+1,s+1} \Sigma_s (S_s^{\mathrm{T}})^{-1} = S_s^{*-1} \phi^*_{s+1,s+1} S_s S_s^{\mathrm{T}} (S_s^{\mathrm{T}})^{-1} = S_s^{*-1} \phi^*_{s+1,s+1} S_s.$$

Therefore, also taking account of (S25), we can write the partial autocorrelation matrices as

$$P_{s+1} = (S_s^{*-1} \phi^*_{s+1,s+1} S_s)^{\mathrm{T}}, \quad (s = 0, \ldots, p-1). \tag{S26}$$

Equations (S24) and (S26) establish the results of step 2(b)(iii) in the forward mapping.

It is also clear that

$$\mathrm{Cov}(\boldsymbol{y}_{t+1}, \boldsymbol{y}_t) = \mathrm{Cov}(\boldsymbol{y}_t, \boldsymbol{y}_{t+1})^{\mathrm{T}} \quad \Longleftrightarrow \quad \phi_{11} \Sigma_0^* = (\phi_{11}^* \Sigma_0)^{\mathrm{T}}$$

and, for $s = 1, \ldots, p-1$, that

$$\mathrm{Cov}(\boldsymbol{\epsilon}_{s,t+1}, \boldsymbol{\epsilon}^*_{s,t-s}) = \mathrm{Cov}(\boldsymbol{\epsilon}^*_{s,t-s}, \boldsymbol{\epsilon}_{s,t+1})^{\mathrm{T}} \quad \Longleftrightarrow \quad \phi_{s+1,s+1} \Sigma_s^* = (\phi^*_{s+1,s+1} \Sigma_s)^{\mathrm{T}}.$$

Therefore we have

$$\phi_{s+1,s+1} \Sigma_s^* = (\phi^*_{s+1,s+1} \Sigma_s)^{\mathrm{T}}, \quad (s = 0, \ldots, p-1). \tag{S27}$$

S1.3 Reverse mapping

In the reverse mapping, the results in step 2(b)(i) from the Appendix of the paper follow from trivial rearrangement of the results in step 2(b)(iii) in the forward mapping. The

equations in step 2(b)(ii) are unchanged from those in step 2(b)(ii) of the forward mapping and the equation in step 2(b)(iv) follows from direct rearrangement of the first equation in step 2(b)(i) of the forward mapping.

We can partition the matrices in the expression for the conditional variance, $\Sigma_{s+1} = \Gamma_0 - \Phi_{s+1}^{\mathrm{T}} g_{s+1}$ $(s = 0, \ldots, p-1)$, as follows

$$\begin{aligned}\Sigma_{s+1} &= \Gamma_0 - \begin{pmatrix} \Phi_{s+1,-(s+1)}^{\mathrm{T}} & \phi_{s+1,s+1} \end{pmatrix} \begin{pmatrix} g_s \\ \Gamma_{s+1} \end{pmatrix} \\ &= \Gamma_0 - \Phi_{s+1,-(s+1)}^{\mathrm{T}} g_s - \phi_{s+1,s+1}\Gamma_{s+1}\end{aligned}$$

which, using (S13), yields

$$\begin{aligned}\Sigma_{s+1} &= \Gamma_0 - (\Phi_s^{\mathrm{T}} - \phi_{s+1,s+1}\Phi_s^{*T} Q) g_s - \phi_{s+1,s+1}\Gamma_{s+1} \\ &= \Gamma_0 - \Phi_s^{\mathrm{T}} g_s - \phi_{s+1,s+1}(\Gamma_{s+1} - \Phi_s^{*T}\tilde{g}_s)\end{aligned}$$

and then using (S2) and (S19) gives

$$\Sigma_{s+1} = \Sigma_s - \phi_{s+1,s+1}\phi_{s+1,s+1}^{*}\Sigma_s. \tag{S28}$$

Finally, from (S27) we have

$$\Sigma_{s+1} = \Sigma_s - \phi_{s+1,s+1}\Sigma_s^{*}\phi_{s+1,s+1}^{\mathrm{T}}, \quad (s = 0, \ldots, p-1). \tag{S29}$$

We can partition the matrices in the expression for the conditional variance, $\Sigma_{s+1}^{*} = \Gamma_0 - \Phi_{s+1}^{*\,T} g_{s+1}^{*}$ $(s = 0, \ldots, p-1\ )$, as follows

$$\begin{aligned}\Sigma_{s+1}^{*} &= \Gamma_0 - \begin{pmatrix} \Phi_{s+1,-(s+1)}^{*\,T} & \phi_{s+1,s+1}^{*} \end{pmatrix} \begin{pmatrix} g_s^{*} \\ \Gamma_{s+1}^{\mathrm{T}} \end{pmatrix} \\ &= \Gamma_0 - \Phi_{s+1,-(s+1)}^{*T} g_s^{*} - \phi_{s+1,s+1}^{*}\Gamma_{s+1}^{\mathrm{T}}\end{aligned}$$

which, using (S15), yields

$$\begin{aligned}\Sigma_{s+1}^{*} &= \Gamma_0 - (\Phi_s^{*T} - \phi_{s+1,s+1}^{*}\Phi_s^{\mathrm{T}} Q) g_s^{*} - \phi_{s+1,s+1}^{*}\Gamma_{s+1}^{\mathrm{T}} \\ &= \Gamma_0 - \Phi_s^{*T} g_s^{*} - \phi_{s+1,s+1}^{*}(\Gamma_{s+1}^{\mathrm{T}} - \Phi_s^{\mathrm{T}}\tilde{g}_s^{*})\end{aligned}$$

and then using (S4) and (S17) gives

$$\Sigma^*_{s+1} = \Sigma^*_s - \phi^*_{s+1,s+1}\phi_{s+1,s+1}\Sigma^*_s.$$

Finally, from (S27) we have

$$\Sigma^*_{s+1} = \Sigma^*_s - \phi^*_{s+1,s+1}\Sigma_s\phi^{*T}_{s+1,s+1}, \quad (s = 0, \ldots, p-1). \tag{S30}$$

Equations (S29) and (S30) establish the results of step 2(b)(iii) in the reverse mapping.

Finally, using the result in step 2(b)(i) of the reverse mapping we can write (S28) as

$$\begin{aligned}
\Sigma_{s+1} &= \Sigma_s - \phi_{s+1,s+1}\phi^*_{s+1,s+1}\Sigma_s \\
&= S_sS_s^{\mathrm{T}} - (S_sP_{s+1}S_s^{*\,-1})(S_s^*P_{s+1}^{\mathrm{T}}S_s^{-1})S_sS_s^{\mathrm{T}} \\
&= S_sS_s^{\mathrm{T}} - S_sP_{s+1}P_{s+1}^{\mathrm{T}}S_s^{\mathrm{T}} \\
&= S_s(I_m - P_{s+1}P_{s+1}^{\mathrm{T}})S_s^{\mathrm{T}}, \quad (s = 0, \ldots, p-1),
\end{aligned} \tag{S31}$$

in which $I_m - P_{s+1}P_{s+1}^{\mathrm{T}}$ is positive semi-definite (Ansley and Kohn, 1986). Equation (S31) establishes the result in step 1(b) of the reverse mapping.

If S_s is the lower triangular Cholesky factor of Σ_s, then we decompose $(I_m - P_{s+1}P_{s+1}^{\mathrm{T}})$ according to its Cholesky decomposition which we write as

$$I_m - P_{s+1}P_{s+1}^{\mathrm{T}} = B_{s+1}^{-1}B_{s+1}^{-1\,T},$$

where B_{s+1}^{-1} is lower triangular, and then solve (S31) for S_s through

$$S_s = S_{s+1}B_{s+1}.$$

If S_s is the symmetric matrix-square-root of Σ_s, then we denote by B_{s+1}^{-1} the symmetric matrix-square-root of $(I_m - P_{s+1}P_{s+1}^{\mathrm{T}})$ and then solve (S31) for S_s through

$$S_s = B_{s+1}(B_{s+1}^{-1}\Sigma_{s+1}B_{s+1}^{-1})^{1/2}B_{s+1}.$$

S2 Invariance to orthogonal transformation

Assume that symmetric matrix-square-roots are used in both parts of the reparameterization of a stationary $\mathrm{VAR}_m(p)$ model for $\boldsymbol{y}_t$ $(t = 1, 2, \ldots)$. In Section 3.2 of the manuscript

we stated that for any $m \times m$ orthogonal matrix H, the parameters of the stationary $\mathrm{VAR}_m(p)$ model for $\tilde{\boldsymbol{y}}_t = H\boldsymbol{y}_t$ are $\tilde{\Sigma} = H\Sigma H^{\mathrm{T}}$ and $\tilde{A}_s = HA_sH^{\mathrm{T}}$ $(s = 1, \ldots, p)$. In this section, we show why this is the case.

The multivariate normal distribution is closed under linear transformation and determined completely by its first two moments. Consider a permutation, rotation or any other orthogonal transformation of the observation vectors, that is $\tilde{\boldsymbol{y}}_t = H\boldsymbol{y}_t$, where H is an $m \times m$ orthogonal matrix. It follows that $\tilde{\boldsymbol{y}}_1, \tilde{\boldsymbol{y}}_2, \ldots$ will follow a stationary $\mathrm{VAR}_m(p)$ process characterized by parameters

$$\{\tilde{\Sigma}, (\tilde{\phi}_1, \ldots, \tilde{\phi}_p)\} \in \mathcal{S}_m^+ \times \mathcal{C}_{p,m} \tag{S32}$$

in which $\tilde{\Sigma} = H\Sigma H^{\mathrm{T}}$ and $\tilde{\phi}_s = H\phi_s H^{\mathrm{T}}$ $(s = 1, \ldots, p)$. Moreover, because H is orthogonal, the symmetric square-roots of $\tilde{\Sigma} = H\Sigma H^{\mathrm{T}}$ and Σ are similar, with $\tilde{\Sigma}^{1/2} = H\Sigma^{1/2}H^{\mathrm{T}}$. Correspondingly, in equation (1) of the manuscript, the coefficients in the conditional expectations for the transformed process in the forward and reverse autoregressions on $s \leq p$ terms are $\tilde{\phi}_{si} = H\phi_{si}H^{\mathrm{T}}$ and $\tilde{\phi}^*_{si} = H\phi^*_{si}H^{\mathrm{T}}$ $(s = 1, \ldots, p;\ i = 1, \ldots, s)$ and the matrix-square-roots of the conditional variance matrices are $\tilde{\Sigma}_s^{1/2} = H\Sigma_s^{1/2}H^{\mathrm{T}}$ and $\tilde{\Sigma}_s^{*\,1/2} = H\Sigma_s^{*\,1/2}H^{\mathrm{T}}$ $(s = 0, \ldots, p)$. The matrix triples $(\tilde{\Sigma}_s^{-1/2}, \tilde{\phi}_{s+1,s+1}, \tilde{\Sigma}_s^{*\,1/2})$ and $(\Sigma_s^{-1/2}, \phi_{s+1,s+1}, \Sigma_s^{*\,1/2})$ are therefore simultaneously similar. Hence, using symmetric matrix-square-roots in (S24), the $(s+1)$th partial autocorrelation for the transformed process is

$$\begin{aligned}
\tilde{P}_{s+1} &= \tilde{\Sigma}_s^{-1/2}\tilde{\phi}_{s+1,s+1}\tilde{\Sigma}_s^{*\,1/2} \\
&= H\Sigma_s^{-1/2}H^{\mathrm{T}}H\phi_{s+1,s+1}H^{\mathrm{T}}H\Sigma_s^{*\,1/2}H^{\mathrm{T}} \\
&= HP_{s+1}H^{\mathrm{T}} \quad (s = 0, \ldots, p-1).
\end{aligned}$$

It then follows trivially from equation (2) in the manuscript that in the unconstrained parameterization, $\tilde{A}_{s+1} = HA_{s+1}H^{\mathrm{T}}$ $(s = 0, \ldots, p-1)$. Hence the parameters (S32) map to $\{\tilde{\Sigma}, (\tilde{P}_1, \ldots, \tilde{P}_p)\} \in \mathcal{S}_m^+ \times \mathcal{V}_m^p$ or, equivalently, $\{\tilde{\Sigma}, (\tilde{A}_1, \ldots, \tilde{A}_p)\} \in \mathcal{S}_m^+ \times M_{m\times m}(\mathbb{R})^p$ where $\tilde{P}_s = HP_sH^{\mathrm{T}}$ and $\tilde{A}_s = HA_sH^{\mathrm{T}}$ $(s = 1, \ldots, p)$. It is important to note that if Cholesky factors are used in the first or second part of the reparameterization, the partial autocorrelation matrices $\tilde{P}_s$ and their real-valued transformations $\tilde{A}_s$ are not similarity

transformations of P_s and A_s. This was one reason for choosing symmetric square-roots, rather than Cholesky factors, in our mappings.

S3 The parameterization of Roy et al. (2019)

S3.1 A parameterization in terms of orthogonal and symmetric positive definite matrices

As discussed in Section 3.5 of the manuscript, based on a characterization of the process in terms of positive definite block Toeplitz matrices, Roy et al. (2019) establish a bijective mapping between the parameters of a stationary $\text{VAR}_m(p)$ process $(\Sigma, \Phi) \in \mathcal{S}_m^+ \times \mathcal{C}_{p,m}$ and the parameter set $\{\Sigma, (V_1, \ldots, V_p), (Q_1, \ldots, Q_p)\} \in \mathcal{S}_m^+ \times \mathcal{S}_m^{+\,p} \times \mathcal{O}(m)^p$ for any fixed choice of a pseudo error variance matrix $M \in \mathcal{S}_m^+$. Unfortunately, for general M, parameter interpretation is difficult. In the special case when $M = \Sigma$, it is straightforward to show that the V_s $(s = 1, \ldots, p)$ are differences between conditional variances $V_s = \Sigma_{s-1} - \Sigma_s$, whilst the Q_s satisfy $Q_s = (\Sigma_{s-1} - \Sigma_s)^{-1/2} \Sigma_{s-1}^{1/2} P_s$, whence $P_s^{\mathrm{T}} \Sigma_{s-1}^{1/2} = Q_s^{\mathrm{T}} (\Sigma_{s-1} - \Sigma_s)^{1/2} = Q_s^{\mathrm{T}} V_s^{1/2}$. As such, Q_s^{T} is the orthogonal matrix arising from the polar decomposition of $P_s^{\mathrm{T}} \Sigma_{s-1}^{1/2}$. The V_s, or at least their diagonal elements, have a clear interpretation as the reduction in residual variance when adding m covariates (the lag-s terms) to a multiple linear regression model. However, even for this special case, interpretation of the orthogonal matrices Q_s is much less clear, which is a significant impediment to prior specification under this parameterization. From a computational perspective, as explained in the manuscript, designing a MCMC scheme that is capable of sampling directly from the space of orthogonal matrices also presents a challenge.

Notwithstanding the problems of interpretation, Roy et al. (2019) attempt to get around the sampling issue through a second reparameterization which maps $\{\Sigma, (V_1, \ldots, V_p), (Q_1, \ldots, Q_p)\}$ to unconstrained Euclidean space. They call this a modified Cayley transform. Though the authors claim the mapping is bijective, it can readily be verified that this is not the case, and this significantly impairs the performance of computational inference by MCMC. Section S3.2 explains and illustrates these points through a series of ex-

amples. Section S3.3 then shows how a bijective mapping of $\{\Sigma, (V_1, \ldots, V_p), (Q_1, \ldots, Q_p)\}$ to uncontrained Euclidean space is available and that a particular choice of distribution for the new parameters provides what might be regarded as a vague prior distribution over the stationary region.

S3.2 The modified Cayley transform is not bijective

In order to avoid designing a MCMC scheme that samples directly from the space of orthogonal matrices, Roy et al. (2019) use a modified Cayley transform to map each $Q_j \in \mathcal{O}(m)$ to a set of unconstrained real-valued parameters $\boldsymbol{s}_j \in \mathbb{R}^{m(m-1)/2}$ which comprise the below diagonal elements of a skew-symmetric matrix S_j, and a binary parameter $\delta_j \in \{0, 1\}$. The symmetric positive definite matrices $V_j \in \mathcal{S}_m^+$ are similarly mapped to Euclidean space through a square-root-free Cholesky factorization (Lindstrom and Bates, 1988). Dropping the lag-j subscript for clarity and denoting $\boldsymbol{e}_1 = (1, 0, \ldots, 0)^{\mathrm{T}}$, the mapping that determines Q from $(\boldsymbol{s}, \delta)$ is given by

$$Q = E_\delta \left\{(I_m - S)(I_m + S)^{-1}\right\}^2 = E_\delta R, \tag{S33}$$

in which $E_\delta = I_m - 2\delta \boldsymbol{e}_1 \boldsymbol{e}_1^T$ is a Householder reflection whose role is to map $Q \in \mathcal{O}(m)$ to $R = \{(I_m - S)(I_m + S)^{-1}\}^2 \in \mathcal{SO}(m)$, where $\mathcal{SO}(m)$ denotes the special orthogonal group, and then $R = \{(I_m - S)(I_m + S)^{-1}\}^2$ parameterizes the special orthogonal group. The inverse mapping that calculates $(\boldsymbol{s}, \delta)$ from Q is described in Section 3.1 of the Supplementary Materials to Roy et al. (2019). First, since R cannot have an odd number of negative one eigenvalues, δ is determined as 0 if Q has an even number of negative one eigenvalues and 1 otherwise, yielding $R = E_\delta Q$. By computing a real square root of R that lies in $\mathcal{SO}(m)$ and does not admit negative one as an eigenvalue, S can then be determined through

$$S = 2\left(I_m + R^{1/2}\right)^{-1} - I_m$$

by using properties of the Cayley transform. The authors claim that this transformation is bijective. However, because a general $R \in \mathcal{SO}(m)$ does not have a *unique* square root with

these properties, it can readily be verified that this is not the case. For example, consider

$$R = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \in \mathcal{SO}(2),$$

where $\theta \in (-\pi, \pi)$. Then both

$$R_A^{1/2} = \begin{pmatrix} \cos(\theta/2) & \sin(\theta/2) \\ -\sin(\theta/2) & \cos(\theta/2) \end{pmatrix} \qquad \text{and} \qquad R_B^{1/2} \begin{pmatrix} -\cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & -\cos(\theta/2) \end{pmatrix}$$

are real square roots of R that lie in $\mathcal{SO}(2)$ and do not admit negative one as an eigenvalue. It can readily be verified that each square root gives a different value of S both of which, when coupled with δ, give the same value of Q through (S33). The mapping from reals to orthogonals is not, therefore, injective.

Suppose a distribution on $\mathbb{R}^{m(m-1)/2} \times \{0, 1\}$ has been specified for $(\boldsymbol{s}, \delta)$. Although it is many-to-one, the above mapping from reals to orthogonals still induces a valid distribution for Q on $\mathcal{O}(m)$. In theory, MCMC methods can therefore be used to draw from the posterior for $(\boldsymbol{s}, \delta)$ and these samples can be transformed through (S33) to give draws from the posterior for Q. In practice, however, the posterior for $(\boldsymbol{s}, \delta)$ will be multimodal which can be a serious impediment to the convergence and mixing of the sampler. Specifically, if the chain is unable to move efficiently between modes, it cannot accurately apportion posterior mass between them. If, in turn, these modes differ in prior support, the approximation of the posterior for Q will not average the likelihood for Q correctly over its prior.

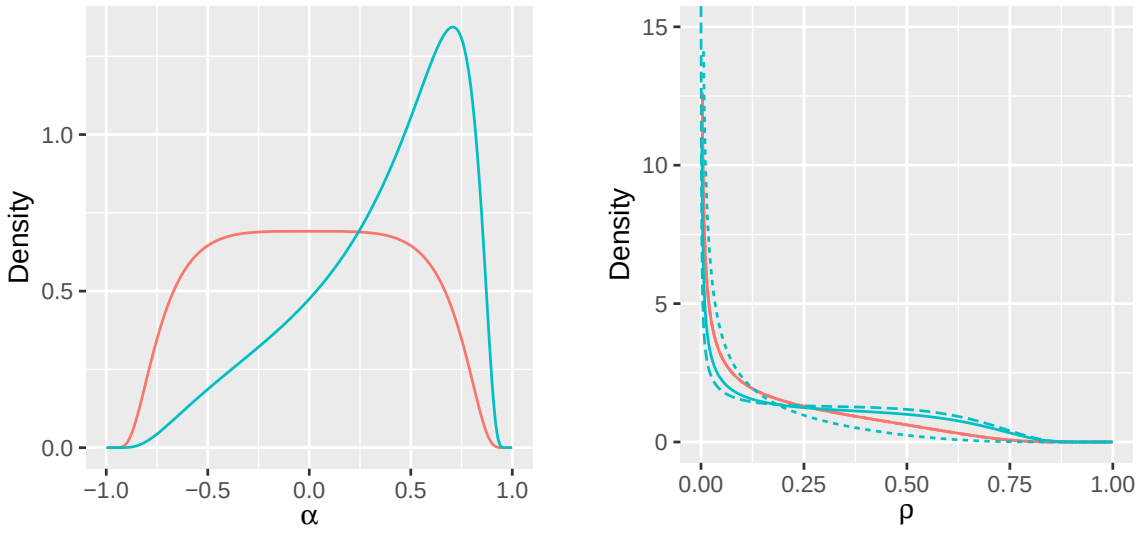
To illustrate the problem, consider the following toy example. Suppose a single observation is to be made on a binomial random variable, $Y|\rho \sim \text{Bin}(n, \rho)$ where $\rho \in (0, 1)$ and $n = 10$. Suppose further that the model is reparameterized in terms of $\alpha = \pm\sqrt{\rho} \in (-1, 1)$ and that α is assigned a prior by taking $\alpha = \zeta(1 + \zeta^2)^{-1/2}$ where $\zeta \sim \text{N}(m, s^2)$. The prior densities induced for α and ρ are

$$\pi_\alpha(\alpha) = \frac{1}{\{2\pi s^2(1-\alpha^2)^3\}^{1/2}} \exp\left[-\frac{1}{2s^2}\left\{\alpha(1-\alpha^2)^{-1/2} - m\right\}^2\right], \qquad -1 < \alpha < 1,$$

and

$$\pi_\rho(\rho) = \frac{1}{2\sqrt{\rho}}\left\{\pi_\alpha(-\sqrt{\rho}) + \pi_\alpha(\sqrt{\rho})\right\}, \qquad 0 \leq \rho < 1.$$

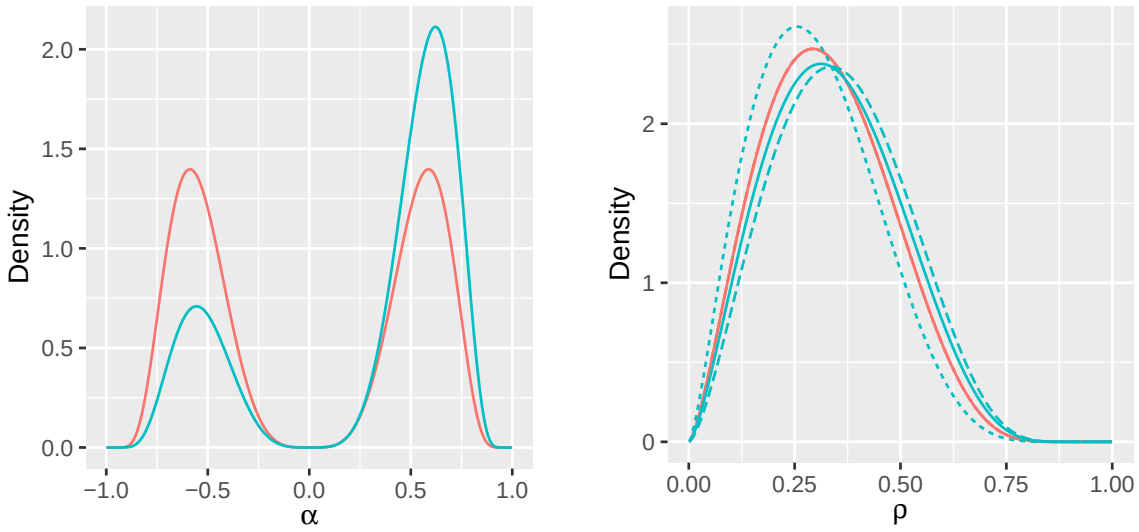
Under two choices for the hyperparameters, $(m = 0, s = \sqrt{3}/3)$ and $(m = 1/4, s = \sqrt{3}/3)$, the prior densities for α and ρ are illustrated in Figure S1. Figure S1(b) also shows the densities for ρ when the distribution for α from which it is generated is truncated at zero on the right and left. The hyperparameter s is chosen to be $\sqrt{3}/3$ because when $m = 0$, the density for α becomes u-shaped when $s > \sqrt{3}/3$.



(a) (b)

Figure S1: Prior density for (a) $\alpha \in (-1, 1)$ and (b) $\rho = \alpha^2 \in [0, 1)$ when $m = 0$ (—) and $m = 0.25$ (—). Also shown in (b) are the prior densities when the distribution for α is truncated to $\alpha \in (-1, 0]$ (· · ·) and $\alpha \in [0, 1)$ (---).

Suppose we go on to observe $Y = y$ where $y \in \{0, 1, \ldots, 10\}$. Irrespective of the value of y, $\Pr(Y = y|\rho = \alpha^2) = \Pr\{Y = y|\rho = (-\alpha)^2\}$ and so the likelihood for α will be bimodal and symmetric about 0. Taking $y = 2$ for illustration and then calculating the marginal likelihoods by numerical integration, the posteriors for α under the two different choices for (m, s) are shown in Figure S2(a). The posterior density clearly inherits the multimodality of the likelihood but it is only symmetric about 0 when the prior is symmetric about 0, that is, when $m = 0$. It is noticeable that for both values of m, the two modes are separated by a region of low posterior density around zero. The corresponding posterior densities for ρ are unimodal and shown in Figure S2(b).



(a) (b)

Figure S2: Posterior density for (a) $\alpha \in (-1, 1)$ and (b) $\rho = \alpha^2 \in [0, 1)$ when $m = 0$ (—) and $m = 0.25$ (—). Also shown in (b) are the posterior densities when the prior distribution for α is truncated to $\alpha \in (-1, 0]$ (· · ·) and $\alpha \in [0, 1)$ (- - -).

If a MCMC sampler was set up to target the posterior for α, it may get stuck in either the negative mode or the positive mode. In this example, the marked separation of the modes is such that a sampler stuck in the negative mode essentially draws α, and hence ρ, from the posterior that would be obtained if the prior for α was truncated on the right at 0. Similarly, a sampler stuck in the positive mode is essentially drawing from the posterior for α, and hence ρ, that would result from truncating the prior for α on the left at 0. These pseudo-posterior densities for ρ are overlaid on the plots in Figure S2(b). When the prior does not offer equal weight to the two values of α that map to the same value of ρ, in this case $\alpha = \pm\sqrt{\rho}$, it is clear that the pseudo-posterior distributions differ from the posterior distribution. Even if the sampler is able to jump between modes in the posterior for α, unless it moves frequently enough to accurately approximate the mass of each mode, the approximation of the posterior for ρ will still be biased.

Unfortunately, these problems manifest clearly when stationary vector autoregressions are reparameterized using the real-valued parameterization of Roy et al. (2019) for problems

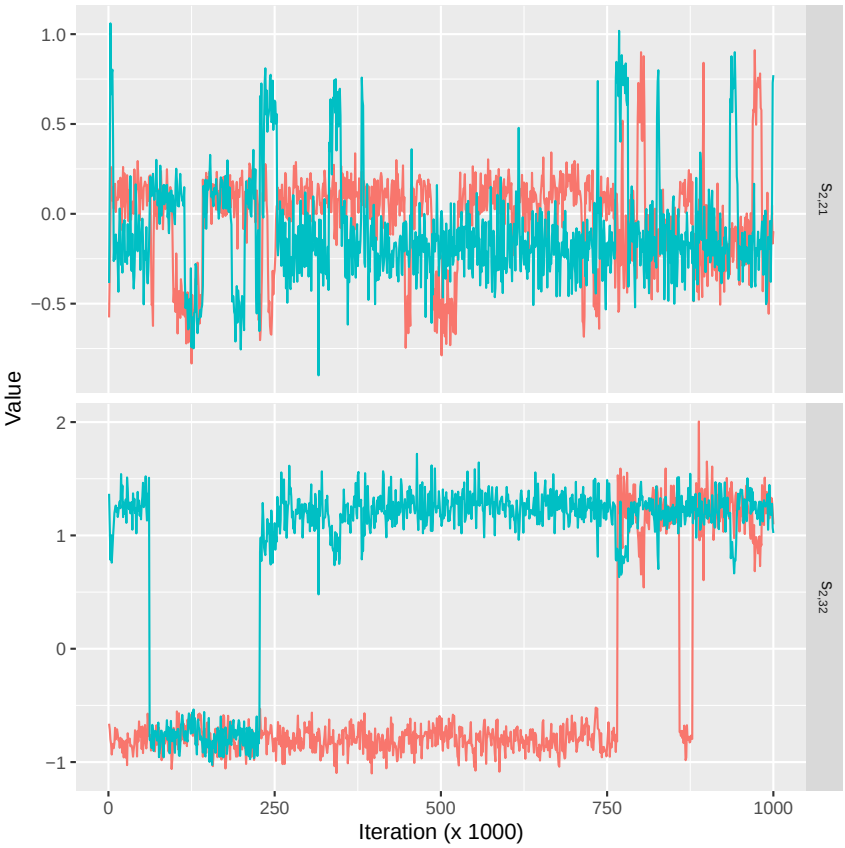


Figure S3: Trace plots obtained under two chains (— and —) for two selected elements in $\boldsymbol{s}_2$. The marginal posteriors are clearly multimodal but jumps between modes are infrequent.

of even moderate complexity. For example, consider the simplest application in Section 5 where $m = 3$ and $p = 4$. We take the prior specification recommended by the authors, which comprises $\mathrm{N}(0, 5)$ distributions for all continuous parameters and a $\mathrm{Bern}(0.5)$ distribution for the reflection parameters. We also use the MCMC algorithm described by the authors, composed of Gaussian random walks for all continuous parameters and an independence sampler for the reflection parameters, with proposal equal to the prior. Two long chains, initialised at different starting points, were run for 1.5M iterations. After omitting the first 500K as burn-in and thinning the remaining draws to retain every 1000th iterate, the trace plots for two elements of the skew-symmetric matrices S_j are shown in Figure S3. It is clear that the posterior is multimodal but the modes are not symmetric about the mean

of the prior (zero) and therefore attract different prior support. In order to accurately approximate the posterior for the elements of the orthogonal matrices Q_j and hence the autoregressive coefficient matrices ϕ_j, it is therefore necessary for the sampler to jump frequently between the modes in the posterior for the S_j. However, this does not happen, and the two chains switch only occasionally between modes, meaning neither will provide a very good approximation to the posterior for (Σ, Φ). Not surprisingly, therefore, the approximations of the marginal posteriors for the parameters of the vector autoregression differ between the two chains; see Figure S4. Although it is possible that this problem could be obviated by doing even longer MCMC runs, this solution is neither practical nor scalable. Indeed, for the more complicated models in Section 5 of the manuscript, when $m = 10$ and $m = 20$, the sampler simply did not converge. Therefore except for simple cases, the unconstrained parameterization and prior of Roy et al. (2019) does not provide a tenable solution to enforcing stationarity through the prior in Bayesian vector autoregressions.

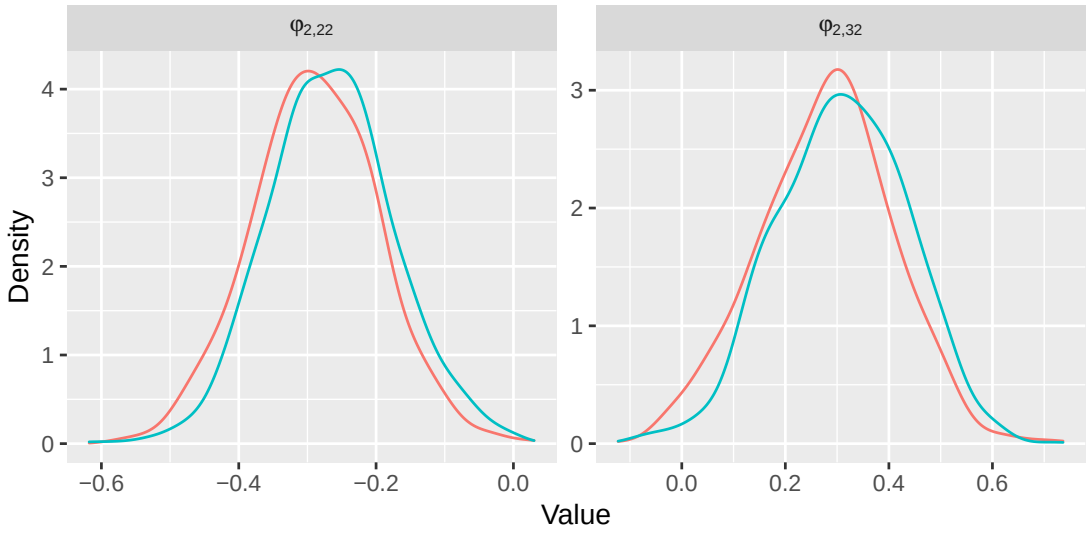


Figure S4: Marginal posterior distribution approximated under two chains (— and —) for two selected elements in ϕ_2. The posterior densities do not overlap, suggesting a lack of convergence in one or both chains.

S3.3 A new reparameterization and vague prior

The unconstrained parameterization discussed in the previous section is problematic because the proposed mapping from reals to orthogonals is not injective. However, it is possible to find a different mapping from $\{\Sigma, (V_1, \ldots, V_p), (Q_1, \ldots, Q_p)\}$ to unconstrained Euclidean space that is a bijection.

Recall from Section S3.1 that in the special case when $M = \Sigma$, the V_s $(s = 1, \ldots, p)$ represent differences between conditional variances whilst the Q_s satisfy $Q_s = (\Sigma_{s-1} - \Sigma_s)^{-1/2}\Sigma_{s-1}^{1/2}P_s$ so that

$$P_s^{\mathrm{T}}\Sigma_{s-1}^{1/2} = Q_s^{\mathrm{T}}(\Sigma_{s-1} - \Sigma_s)^{1/2} = Q_s^{\mathrm{T}}V_s^{1/2} \tag{S34}$$

and hence Q_s^{T} is the orthogonal matrix arising from the polar decomposition of $P_s^{\mathrm{T}}\Sigma_{s-1}^{1/2}$. The polar decomposition is a unique representation of a full rank, real-valued matrix as the product of an orthogonal matrix and a symmetric, positive definite matrix. Therefore since the parameter space of Q_s^{T} is $\mathcal{O}(m)$ and that of V_s is $\mathcal{S}_m^+$, it follows from (S34) that the parameter space of $P_s^{\mathrm{T}}\Sigma_{s-1}^{1/2}$ is $M_{m\times m}(\mathbb{R})$. Writing $C_s = P_s^{\mathrm{T}}\Sigma_{s-1}^{1/2} = Q_s^{\mathrm{T}}V_s^{1/2}$, we therefore have another bijection between $(\Sigma, \Phi) \in \mathcal{S}_m^+ \times \mathcal{C}_{p,m}$ and $\{\Sigma, (C_1, \ldots, C_p)\} \in \mathcal{S}_m^+ \times M_{m\times m}(\mathbb{R})^p$. Although the complex interpretation of the C_s precludes the kind of structural prior specification available for the transformed partial autocorrelations, it is well known (for example, see Jauch et al., 2021; Eaton, 1989, Chapter 5) that if C_s comprises m^2 independent standard normal random variables, then the joint distribution induced for the components of its polar decomposition are such that Q_s and V_s are independent, with Q_s uniformly distributed over $\mathcal{O}(m)$, and V_s Wishart distributed with m degrees of freedom and identity scale. As a corollary to the results from Section 3.2, it is straightforward to show that if an orthogonal transformation is applied to the observations vectors $\tilde{\boldsymbol{y}}_t = H\boldsymbol{y}_t$, then the new parameters that result from this transformation are $\tilde{C}_s = HC_sH^{\mathrm{T}}$. As the standard matrix normal distribution is rotatable (Dawid, 1981), the prior distribution for the C_s will be unchanged by a permutation of the observations $\boldsymbol{y}_t$. It therefore serves as a vague, stationary prior distribution that is additionally exchangeable, whilst allowing the inferential problem to be cast in Euclidean space. This might be attractive to some modellers as a default choice of prior.

S4 Choice of prior variance for the unconstrained square matrices

For the simplest vector autoregression, which is a $\text{VAR}_2(1)$ model, we consider two versions of the exchangeable, stationary prior presented in Section 3.2 of the manuscript. Specifically, we consider the hierarchical prior expressed through equations (7)–(9), with hyperparameters chosen as

$$\begin{aligned}
&\text{Prior 1:} \quad e_{si} = 0, \quad f_{si} = \sqrt{0.35}, \quad g_{si} = 1.05, \quad h_{si} = 0.0075, \\
&\text{Prior 2:} \quad e_{si} = 0, \quad f_{si} = \sqrt{3.5}, \quad\ \ g_{si} = 10.5, \quad h_{si} = 14.25,
\end{aligned}$$

for $s = p = 1$ and $i = 1, 2$. This gives marginal prior means and correlations of $E(a_{1,ii}) = E(a_{1,ij}) = 0.0$ and $\text{Cor}(a_{1,11}, a_{1,22}) = \text{Cor}(a_{1,12}, a_{1,21}) = 0.7$, and marginal prior standard deviations of $\text{SD}(a_{1,ii}) = \text{SD}(a_{1,ij}) = 0.5$ or $\text{SD}(a_{1,ii}) = \text{SD}(a_{1,ij}) = 5.0$, respectively. For comparative purposes, we also consider the vague, stationary prior based on the parameterization of Roy et al. (2019) that was discussed in Section 3.5 of the paper and Section S3.3 above. We refer to this as Prior 3. Since prior beliefs are arguably most naturally expressed in terms of the partial autocorrelation matrices, we visualize the distribution for the elements of P_1 in Figure S5 from which it is clear that in the priors with larger variance, namely Priors 2 and 3, the marginal distributions for all elements are multimodal. The corresponding plot for the elements of the autoregressive coefficient matrix ϕ_1 is shown in Figure S6 under the prior $\Sigma \sim \text{IW}(m+4, \boldsymbol{I}_m)$ for the error variance matrix. In addition to displaying the complex geometry of the stationary region $\mathcal{C}_{1,2}$, this plot reveals that multimodality can also become a feature of the prior under the original (Φ, Σ)-parameterization of the model when the prior variance for the A_s becomes too large. For most problems, a multimodal prior for a partial autocorrelation matrix P_s (or an autoregressive coefficient matrix ϕ_s) is unlikely to be representative of prior beliefs. To avoid this, care is clearly needed in the choice of prior variance for the $a_{s,ij}$.

As explained in Section 2.2 of the manuscript, each unconstrained square matrix A_s is related to the corresponding partial autocorrelation matrix P_s by a simple mapping of the singular values from the positive real line to the unit interval. It is reasonable,

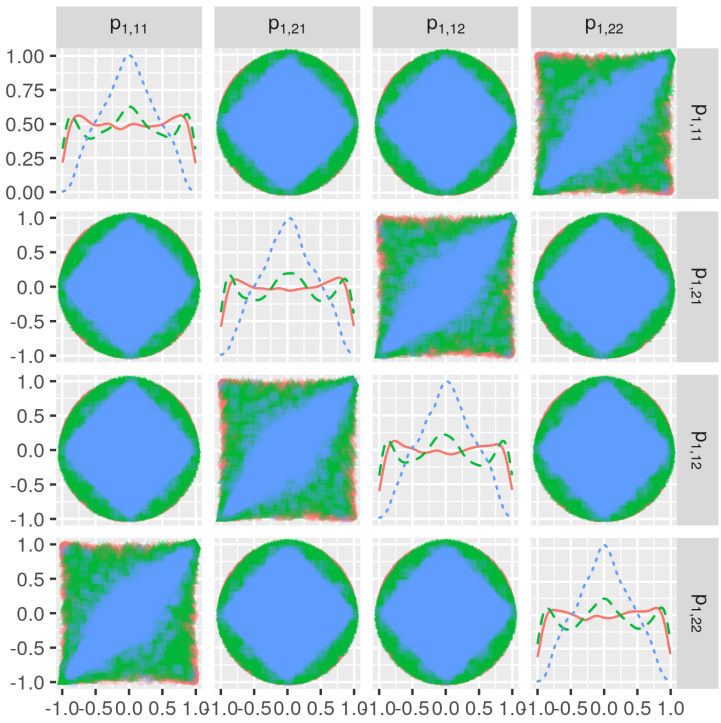


Figure S5: Visualization of the prior induced for P_1 in a $\text{VAR}_2(1)$ model: Priors 1 (□, · · ·); 2 (△, ---); 3 (○, ——). All plots share a common x-scale. Diagonal plots depict marginal densities and share a common y-scale (see the y-axis of the $(1,1)$ plot). Off-diagonal plots depict bivariate densities and share a common y-scale (see the y-axis of the $(j,1)$ plots for $j = 2,3,4$).

therefore, to conjecture that the multimodality that can occur in the prior for the partial autocorrelations, but not in the multivariate normal prior for the unconstrained square matrices, arises through this mapping of the singular values. Dropping the lag-s subscript for brevity, we can gain insight into the behaviour of the prior induced for the singular values, along with the right and left singular vectors, of P through a closed form expression for their joint density in the special case when all the elements of $A = (a_{ij})$ are independent and identically distributed *a priori*, with $a_{ij} \sim \text{N}(0, s^2)$ $(i, j = 1, \ldots, m)$.

Denote the singular value decomposition of A by $A = U\tilde{R}V^{\text{T}}$ where $\tilde{R}$ is a diagonal matrix whose diagonal values are the singular values $\tilde{r}_1, \ldots, \tilde{r}_m$ of A and where U and V are its left and right singular vectors, respectively. If the singular values are ordered,

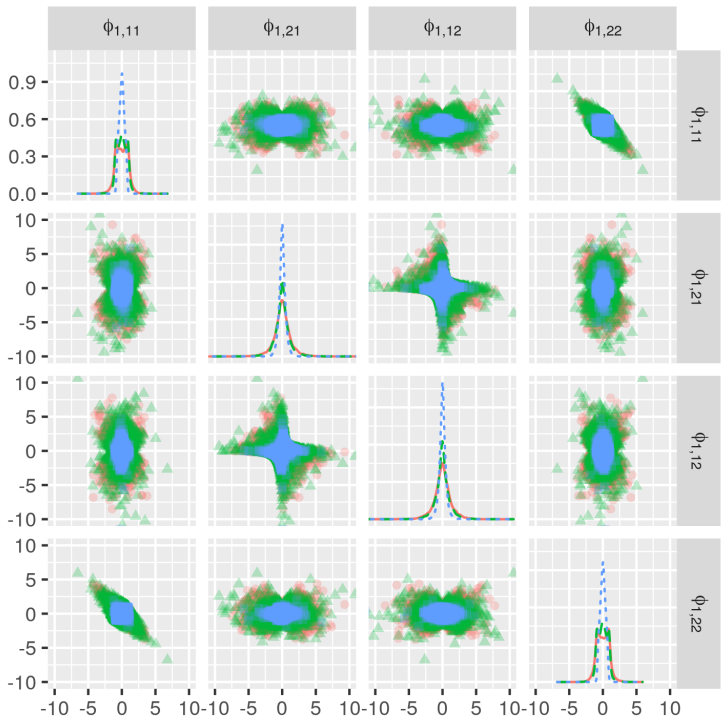


Figure S6: Visualization of the prior induced for ϕ_1 in a $\text{VAR}_2(1)$ model: Priors 1 (□, · · ·); 2 (△, ---); 3 (○, ——). All plots share a common x-scale. Diagonal plots depict marginal densities and share a common y-scale (see the y-axis of the $(1,1)$ plot). Off-diagonal plots depict bivariate densities and share a common y-scale (see the y-axis of the $(j,1)$ plots for $j = 2,3,4$).

distinct and positive and we fix the sign of the m right (or left) singular vectors, then this transformation is unique. Using the Jacobian of the singular value decomposition (Edelman and Rao, 2005) it is straightforward to derive the prior distribution of U, V and the $\tilde{r}_i$ under these conditions through

$$\pi(A)(dA) = (2\pi)^{-m^2/2} s^{-m^2} \exp\left(-\frac{1}{2s^2}\sum_{i=1}^{m}\tilde{r}_i^2\right)\prod_{i=1}^{m-1}\prod_{j=i+1}^{m}(\tilde{r}_i^2 - \tilde{r}_j^2)(U^{\mathrm{T}}dU)(d\tilde{R})(V^{\mathrm{T}}dV).$$

It follows that U, V and $(\tilde{r}_1, \ldots, \tilde{r}_m)$ are independent *a priori* and that the distributions of U and V are normalized Haar measures, the latter restricted to one pattern of column signs. Further, we can derive the marginal distribution of the singular values $\tilde{r}_1, \ldots, \tilde{r}_m$ of

A as

$$\pi(\tilde{r}_1, \ldots, \tilde{r}_m) = \frac{\pi^{m^2/2}}{2^{m(m-2)/2} s^{m^2} \Gamma_m(m/2)^2} \exp\left(-\frac{1}{2s^2} \sum_{i=1}^{m} \tilde{r}_i^2\right) \prod_{i=1}^{m-1} \prod_{j=i+1}^{m} (\tilde{r}_i^2 - \tilde{r}_j^2),$$

for $\tilde{r}_1 > \tilde{r}_2 > \cdots > \tilde{r}_m > 0$ where $\Gamma_m(x) = \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma\{x - (i-1)/2\}$ is the multivariate gamma function. The singular value decomposition of P is simply $P = URV^{\mathrm{T}}$ in which the diagonal matrix has ith diagonal entry $r_i = \tilde{r}_i(1+\tilde{r}_i^2)^{-1/2}$ $(i = 1, \ldots, m)$. In the prior induced for the singular values and vectors of P, the singular values therefore remain independent of the singular vectors and the singular vectors remain distributed according to independent normalized Haar measures. The Jacobian J of the transformation from the singular values of A to those of P has determinant

$$|J| = \prod_{i=1}^{m} (1 - r_i^2)^{-3/2},$$

from which it follows that the joint density for the singular values of P is

$$\pi(r_1, \ldots, r_m) = \frac{\pi^{m^2/2}}{2^{m(m-2)/2} s^{m^2} \Gamma_m(m/2)^2} \exp\left(-\frac{1}{2s^2} \sum_{i=1}^{m} \frac{r_i^2}{1 - r_i^2}\right) (1 - r_m^2)^{-(2m+1)/2} \times \prod_{i=1}^{m-1} (1 - r_i^2)^{-(2m+1)/2} \prod_{j=i+1}^{m} (r_i^2 - r_j^2),$$

for $1 > r_1 > r_2 > \cdots > r_m > 0$, where it is understood that the products over i and j evaluate to 1 when $m = 1$. In the univariate case, when $m = 1$, there is a scalar-valued partial autocorrelation, $p = urv$, with $u \sim \mathrm{U}\{-1, 1\}$ and $v = 1$. When the single singular value r has a prior density with a local maximum at, say $r_0 \in (0, 1)$, the prior density for $p = \pm r$ will clearly be bimodal with symmetric modes at $p = -r_0$ and $p = r_0$. It is therefore reasonable to posit that this behaviour will generalize to the multivariate case when $m > 1$. In other words, multimodalities, like those seen in Figure S5, are induced in the prior for P when the joint density of the singular values has a local maximum in $1 > r_1 > r_2 > \cdots > r_m > 0$. For the low-dimensional cases for which visualization of the joint density is feasible, this conjecture is supported by simulation from the prior; see Figures S7 and S8. At the very least, identifying the value of s at which a local maximum first occurs serves as guide for making a principled choice of the variance in the prior for A.

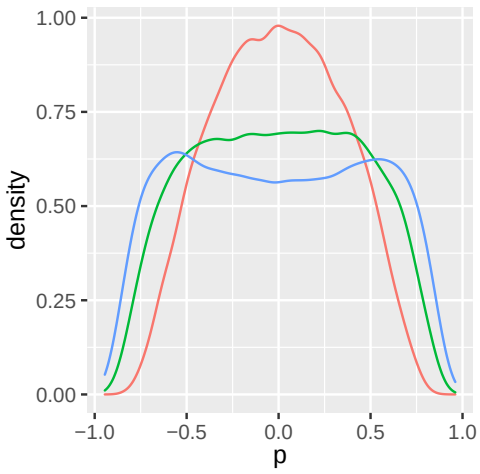


Figure S7: Prior densities induced for the scalar partial autocorrelation p when $m = 1$ and: $s^2 = 1/6$ (—); $s^2 = 1/3$ (—); $s^2 = 1/2$ (—). The density becomes bimodal when $s^2 > 1/3$.

Denoting $\boldsymbol{r} = (r_1, \ldots, r_m)^{\mathrm{T}}$, the components of the gradient of the logarithm of the prior density for $\boldsymbol{r}$ are given by

$$\frac{\partial \log\{\pi(\boldsymbol{r})\}}{\partial r_i} = \frac{r_i \left\{(2m+1)s^2(1-r_i^2) - 1\right\} \prod_{j \neq i}(r_i^2 - r_j^2) + 2s^2 r_i (1-r_i^2)^2 \sum_{j \neq i} \prod_{k \neq i,j}(r_i^2 - r_k^2)}{s^2(1-r_i^2)^2 \prod_{j \neq i}(r_i^2 - r_k^2)},$$

for $i = 1, \ldots, m$, whilst the terms in the associated Hessian matrix are

$$\frac{\partial^2 \log\{\pi(\boldsymbol{r})\}}{\partial r_i^2} = \frac{(2m+1)(1+r_i^2)}{(1-r_i^2)^2} - \frac{1+3r_i^2}{s^2(1-r_i^2)^3} - 2\sum_{j \neq i} \frac{r_i^2 + r_j^2}{(r_i^2 - r_j^2)^2}, \quad (i = 1, \ldots, m),$$

and

$$\frac{\partial^2 \log\{\pi(\boldsymbol{r})\}}{\partial r_i \partial r_j} = \frac{4 r_i r_j}{(r_i^2 - r_j^2)^2}, \quad (i \neq j).$$

Finding stable points in the region $1 > r_1 > r_2 > \cdots > r_m > 0$, if they exist, is therefore tantamount to finding the real roots of the system of m polynomial equations

$$0 = \left\{(2m+1)s^2(1-r_i^2) - 1\right\} \prod_{j \neq i}(r_i^2 - r_j^2) + 2s^2(1-r_i^2)^2 \sum_{j \neq i} \prod_{k \neq i,j}(r_i^2 - r_k^2), \quad (i = 1, \ldots, m),$$

subject to this constraint. For any fixed value of the prior standard deviation s, these equations can, in principle, be solved. In combination with application of the second derivative test, this allows determination of the smallest value of s at which the distribution

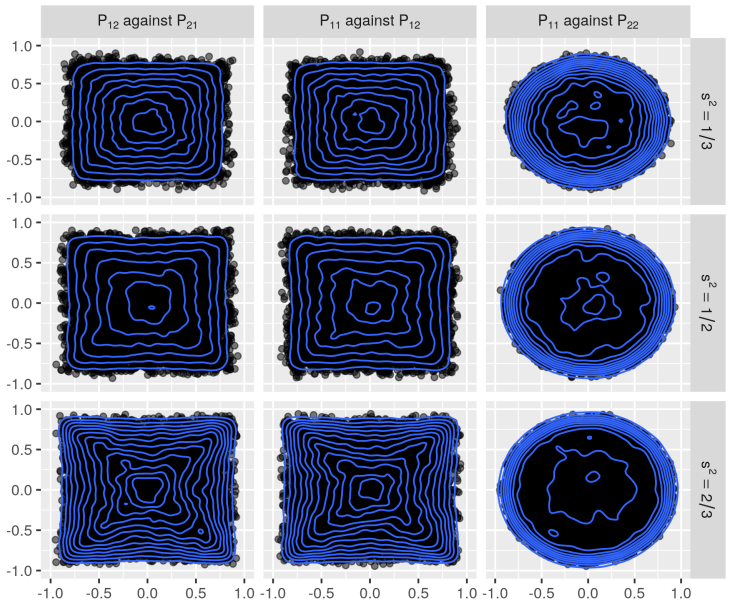


Figure S8: Pairwise densities between a pair of off-diagonal elements (left panels), a pair of diagonal elements (middle panels) and a diagonal and off-diagonal element (right panels) in the joint prior induced for the partial autocorrelation matrix P when $m = 2$ and $s^2 = 1/3$ (top row); $s^2 = 1/2$ (middle row); $s^2 = 2/3$ (bottom row). The density appears to be multimodal for $s^2 > 1/2$.

has a local maximum in $1 > r_1 > r_2 > \cdots > r_m > 0$. For $m = 1$ and $m = 2$, the roots can be found by hand, along with the value of s at which a local maximum first occurs. For $m = 3$, $m = 4$ and $m = 5$ the system can be solved with the help of symbolic computation software such as Maple over a grid of values for s. This allows identification of the approximate value for s where a local maximum first appears. The results are displayed in Table S1. The number of possible solutions to an (unconstrained) system of polynomials, each of degree d, grows exponentially with d and, in this case, is $d = 2m$. Although this can be reduced to $d = m$ by using the substitution $w_i = 1 - r_i^2$, the search still becomes computationally infeasible for $m > 5$. However, from Table S1 it seems that the value of s at which a local maximum first appears grows with m, but at a decreasing rate, such that for $m \geq 5$, a value of $s = 1$ would be a reasonable choice. For smaller values of m, the choice of s can be guided by Table S1.

Table S1: The standard deviation s_0 in the prior for the elements of A such that when $s > s_0$, the prior induced for the singular values $r_1, \ldots, r_m$ of P has a local maximum in $1 > r_m > \cdots > r_1 > 0$. Values for $m \leq 2$ are exact; values for $m > 2$ lie in the stated open interval.

m	1	2	3	4	5
s_0	$1/\sqrt{3}$	$1/\sqrt{2}$	$(0.81, 0.82)$	$(0.90, 0.91)$	$(0.98, 0.99)$

S5 Effect of an ill-conditioned variance matrix on inference

One of the features of the reparameterization for univariate autoregressive models that is not preserved in the vector generalization is the independence of the error variance Σ and partial autocorrelations $P_1, \ldots, P_p$. Indeed, in the vector case, $P_1, \ldots, P_p$ and hence the unconstrained square matrices $A_1, \ldots, A_p$, depend on Σ as well as the autoregressive coefficient matrices $\phi_1, \ldots, \phi_p$. This suggests that if the likelihood is formulated in terms of the new parameters, the dependence of $A_1, \ldots, A_p$ on Σ may distort the shape of the likelihood, thereby causing numerical problems in sampling from the posterior. This, in turn, would cause numerical problems in the approximation of the posterior for $\phi_1, \ldots, \phi_p$.

If such a problem was to arise, it is likely that it would be exacerbated if Σ was close to being singular. In order to investigate whether there is any evidence to substantiate these concerns, we performed a series of simulation experiments. Six data sets of length 1,000 were sampled from a $\text{VAR}_3(2)$ model with error variance (and correlation) matrix

$$\Sigma = \begin{pmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{pmatrix},$$

where $r \in \{0.000, 0.500, 0.833, 0.955, 0.988, 0.997\}$ to three decimal places. The condition number for a matrix of this form is $c = (1 + 2r)/(1 - r)$ and so the matrices had condition numbers $c \in \{1, 4, 16, 64, 256, 1024\}$. For each of the six data sets, common values for the

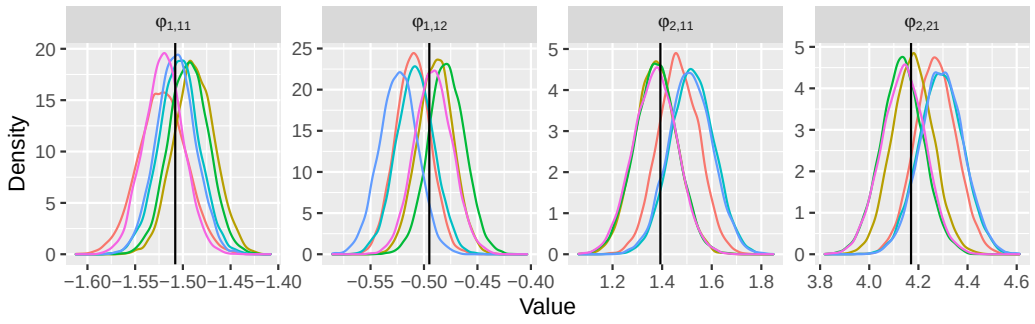
autoregressive coefficient matrices ϕ_1 and ϕ_2 were used. This process was repeated three times using three different pairs of matrices ϕ_1 and ϕ_2, as detailed in Table S2.

Table S2: Autoregressive coefficient matrices, ϕ_1 and ϕ_2, used to simulate the six data sets in each of the three experiments. Also shown for each experiment is the minimum effective sample size when analysing the data simulated using the error variance matrix Σ with each of the six possible values of the condition number c.

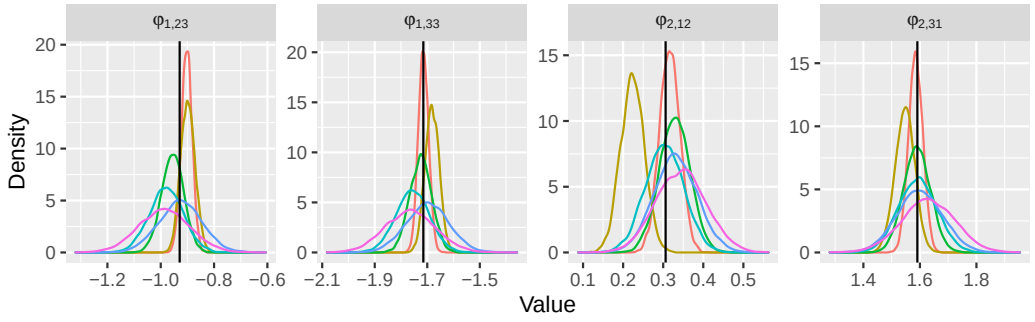
Experiment	ϕ_1	ϕ_2	Minimum Effective Sample Size by c					
			1	4	16	64	256	1024
1	$\begin{pmatrix} -1.508 & -0.495 & -2.500 \\ 4.893 & -0.975 & 0.895 \\ 1.123 & 0.613 & 2.475 \end{pmatrix}$	$\begin{pmatrix} 1.393 & 0.063 & 1.477 \\ 4.170 & 0.204 & 3.261 \\ -2.120 & -0.031 & -1.713 \end{pmatrix}$	3761	2762	2531	3102	2567	3313
2	$\begin{pmatrix} 0.226 & 0.509 & 0.083 \\ 0.499 & -0.195 & -0.929 \\ 1.720 & 1.353 & -1.715 \end{pmatrix}$	$\begin{pmatrix} 1.001 & 0.306 & -0.360 \\ 0.395 & 0.399 & -0.221 \\ 1.591 & 0.932 & -0.810 \end{pmatrix}$	1954	2869	3140	2551	2279	1949
3	$\begin{pmatrix} 0.223 & 0.100 & 1.191 \\ -0.344 & -0.157 & -1.627 \\ 0.531 & -0.078 & 1.118 \end{pmatrix}$	$\begin{pmatrix} -0.502 & 0.340 & -0.874 \\ 0.829 & -0.346 & 0.790 \\ 0.000 & -0.207 & -0.446 \end{pmatrix}$	2874	1769	2196	2634	1952	2065

The unconstrained square matrices A_1 and A_2 and the error variance matrix Σ were given the exchangeable, stationary prior referred to as Prior 1 in Section S4. In order to fit a $\text{VAR}_3(2)$ model to each data set, we used Hamiltonian Monte Carlo, implemented in Stan. In all cases we used the `rstan` interface to the Stan software to run four chains, initialized at different starting points, for 5000 iterations, half of which were discarded as burn-in. After pooling the output across chains, the minimum effective sample sizes from each analysis are shown in Table S2. For a randomly chosen selection of parameters in ϕ_1 and ϕ_2, Figure S9 shows plots of the marginal posteriors based on data simulated in each of the three experiments.

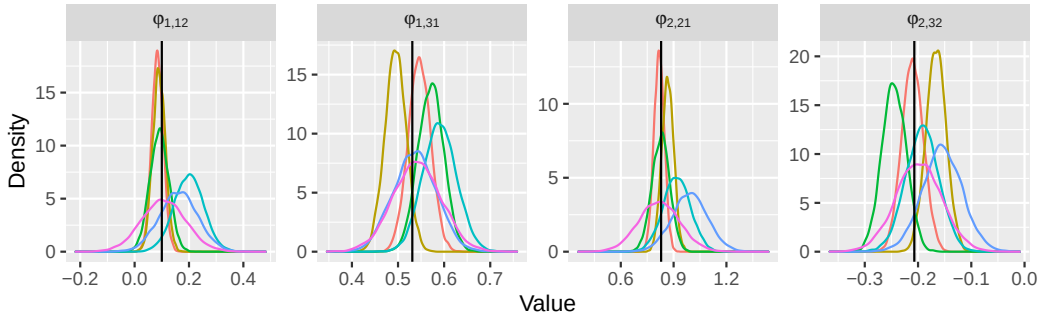
In all the panels of Figure S9, the posteriors are roughly centred at the value used to simulate the data, though it is interesting to note that under some values for ϕ_1 and ϕ_2, such



(a)



(b)



(c)

Figure S9: Marginal posterior distributions for four randomly selected parameters in ϕ_1 and ϕ_2 for the data simulated under each of the six values for Σ in Experiments (a) 1; (b) 2; (c) 3. The values of the parameters used to simulate the data are indicated by solid vertical lines. The condition numbers of the six values for Σ are 1 (—); 4 (—); 16 (—); 64 (—); 256 (—); 1024 (—).

as those used in Experiments 2 and 3, the posterior for ϕ_1 and ϕ_2 becomes more diffuse as the condition number of Σ increases. Crucially, the usual graphical and numerical diagnostics gave no evidence of any lack of convergence and from Table S2, there does not seem to be any deterioration in mixing performance as Σ approaches a singular matrix. Therefore, on the basis of this simulation experiment, there is no evidence that the intertwining of the two parameter sets, $(A_1, \ldots, A_p)$ and Σ, causes any problems in the numerical approximation of the posterior.

S6 Further details of the application

S6.1 Data

Table S3 indicates which variables in the macroeconomic time series analysed in the manuscript were used in the $\text{VAR}_3(4)$, $\text{VAR}_{10}(4)$ and $\text{VAR}_{20}(4)$ models.

S6.2 Prior specification

In the exchangeable, stationary prior used in the application, we use the guidelines from Section S4 to choose the hyperparameters in equations (7)–(9) from the manuscript as

$$\begin{aligned} m = 3: &\qquad e_{si} = 0, \quad f_{si} = \sqrt{0.455}, \quad g_{si} = 1.365, \quad h_{si} = 0.071, \\ m = 10, 20: &\quad e_{si} = 0, \quad f_{si} = \sqrt{0.700}, \quad g_{si} = 2.100, \quad h_{si} = 0.333, \end{aligned}$$

for $s = 1, \ldots, 4$ and $i = 1, 2$. This gives marginal prior means and correlations of $E(a_{s,ii}) = E(a_{s,ij}) = 0.000$ and $\text{Cor}(a_{s,ii}, a_{s,jj}) = \text{Cor}(a_{s,ij}, a_{s,i'j'}) = 0.700$, and marginal prior standard deviations of $\text{SD}(a_{s,ii}) = \text{SD}(a_{s,ij}) = 0.806$ ($m = 3$) or $\text{SD}(a_{1,ii}) = \text{SD}(a_{1,ij}) = 1.000$ ($m = 10$ and $m = 20$).

In the Minnesota and semi-conjugate priors used in the application, the prior means in $\boldsymbol{u}_{k2}$ and $\boldsymbol{u}_{k3}$ ($k = 1, \ldots, p$), were chosen to be zero in all analyses and every element of Φ was taken to be independent, so that the prior variance matrices W_{k2} and W_{k3} were diagonal. The diagonal elements in W_{k2} and W_{k3} were chosen according to the default choices from the model-fitting software provided with the monograph of Koop and Korobilis

Table S3: The variables in the macroeconomic time series and the models (m) in which they were used.

Variable	m
Real GDP: quantity index (2000 = 100)	$3, 10, 20$
CPI: all items	$3, 10, 20$
Interest rate: federal funds (e_ective) (percentage per annum)	$3, 10, 20$
Real spot market price index: all commodities	$10, 20$
Depository institution reserves: nonborrowed (millions of dollars)	$10, 20$
Depository institution reserves: total (millions of dollars)	$10, 20$
Money stock: M2 (billions of dollars)	$10, 20$
Real Personal Consumption Expenditures: quantity index	$10, 20$
Industrial production index: total	$10, 20$
Capacity utilization: manufacturing (SIC)	$10, 20$
Unemployment rate: all workers, 16 and over (percentage)	20
Housing starts: total (thousands)	20
Producer price index: finished goods	20
Personal Consumption Expenditures: price index	20
Real average hourly earnings: non-farm production workers	20
Money stock: M1 (billions of dollars)	20
S&P's common stock price index: industrials	20
Interest rate: US treasury constant maturity, 10-year	20
US effective exchange rate: index number	20
Employees, non-farm: total private	20

(2009) where the same data are analysed using vector autoregressive models. Specifically, in the semi-conjugate prior, $W_{k3} = 10I_{m^2}$ $(k = 1, \ldots, p)$ and in the Minnesota prior, the diagonal elements in the prior variance matrix $\mathrm{Var}\{\mathrm{vec}(\phi_k)\} = W_{k2}$ were chosen such that $\mathrm{Var}(\phi_{k,ii}) = c/k^2$ $(i = 1, \ldots, m)$ and $\mathrm{Var}(\phi_{k,ij}) = ds_i^2/(k^2 s_j^2)$ $(i \neq j)$ with $c = d = 1/2$. In these expressions, s_i^2 is the ordinary least squares estimate of the error variance in the (univariate) autoregression for variable i. The idea is that the prior variance decreases with lag to encourage shrinkage of the autocovariance matrices at higher lag towards zero.

S6.3 Assessment of forecasting performance

In Section 5 of the manuscript, the strategy for assessing forecasting performance is explained. In brief, the posterior (or maximum likelihood estimates) for the model parameters were obtained using the data up to time $n = 156$, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$, and then forecasting performance was assessed using the remaining 40 hold-out observations, $\boldsymbol{y}_{n+1}, \ldots, \boldsymbol{y}_{n+40}$. For a variety of horizons, $h = 1, 2, 4, 8$, the h-step ahead forecasts arising from the four model-prior combinations, along with the model fitted using maximum likelihood, were compared using the MSFE and a number of proper scoring rules.

Given data, $\boldsymbol{y}_{1:n}$, the h-step ahead MSFE for variable k is defined by

$$\mathrm{MSFE}_{h,k}(\boldsymbol{y}_{(n+1):(n+40)}, \Phi) = \frac{1}{40 - h + 1} \sum_{t=n+h}^{n+40} \left[y_{tk} - E\left\{ Y_{tk} | \boldsymbol{y}_{1:(t-h)}, \Phi \right\} \right]^2$$

For each value of $h \in \{1, 2, 4, 8\}$ and $m \in \{3, 10, 20\}$ and each variable of interest $k \in \{1, 2, 3\}$, the posterior distribution for $\log \mathrm{MSFE}_{h,k}(\boldsymbol{y}_{(n+1):(n+40)}, \Phi)$ for the four Bayesian analyses is summarized in Figure S10 through its mean and credible intervals which extend to two posterior standard deviations either side of the mean. For the maximum likelihood analysis, we present the maximum likelihood estimate (MLE) for $\log \mathrm{MSFE}_{h,k}(\boldsymbol{y}_{(n+1):(n+40)}, \Phi)$, namely $\log \mathrm{MSFE}_{h,k}(\boldsymbol{y}_{(n+1):(n+40)}, \hat{\Phi})$ where $\hat{\Phi}$ is the MLE of Φ which can be computed from the MLEs $(\hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma})$ of the parameters in the fitted model. Uncertainty in the value of the estimator is quantified by applying the delta method to an asymptotic approximation to the standard error of the MSFE obtained using the approach of Harvey et al. (1997).

Proper scoring rules assign a numerical score $S(F, y)$ by comparing a forecast distribu-

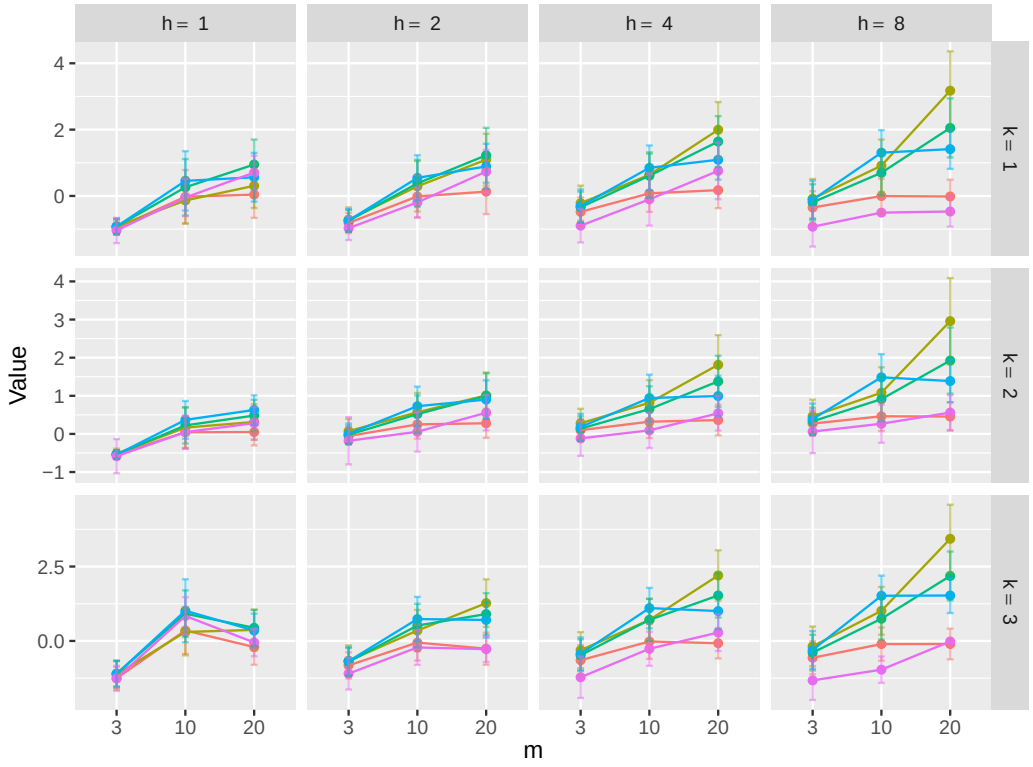


Figure S10: For each value of m, each forecast horizon h, and each prior: posterior mean (with arrows extending to plus or minus two posterior standard deviations) for the logarithm of the empirical MSFE for each variable of interest k. The priors are: exchangeable and stationary (●); Minnesota (●); semi-conjugate (●); vague and stationary (●). Also shown are analogous summaries based on the stationary MLE (Ansley and Kohn, 1986) (●).

tion function F (or density function f) with the observation y that arises. For assessing individual forecasts of the three variables of interest we considered two popular scoring rules: the continuous rank probability score (CRPS) and the logarithmic score. Given data, $\boldsymbol{y}_{1:n}$, the h-step ahead CRPS for variable k at time $t = n + h, \ldots, n + 40$ is defined through

$$\text{CRPS}(F_{h,tk}, y_{tk}) = \int \{F_{h,tk}(z) - \mathbb{I}(y_{tk} \leq z)\}^2 dz = E|Y_{tk} - y_{tk}| - \frac{1}{2}E|Y_{tk} - Y'_{tk}|.$$

Here $\mathbb{I}(x)$ denotes the indicator function, which is equal to 1 if x is true and 0 otherwise, and

Y_{tk}, Y'_{tk} are independent and identically distributed random variables whose h-step ahead predictive distribution function is $F_{h,tk}$. The logarithmic score is defined through

$$\text{logS}(F_{h,tk}, y_{tk}) = -\log f_{h,tk}(y_{tk}),$$

in which $f_{h,tk}$ is the density function associated with $F_{h,tk}$.

We additionally compared the model-prior combinations in terms of their ability to produce joint forecasts by computing the energy score (ES), which is a multivariate generalization of the CRPS. In its most widely used form, the h-step ahead ES is

$$\text{ES}(F_{h,t}, \boldsymbol{y}_t) = E\|\boldsymbol{Y}_t - \boldsymbol{y}_t\| - \frac{1}{2}E\|\boldsymbol{Y}_t - \boldsymbol{Y}'_t\|, \quad t = n+h, \ldots, n+40$$

in which $\boldsymbol{Y}_t, \boldsymbol{Y}'_t$ are independent and identically distributed random vectors whose h-step ahead joint predictive distribution is $F_{h,t}$. We calculated the energy score for a three-dimensional vector containing the variables of interest.

To approximate the scoring rules at each time point t, we used the `scoringRules` package in R (Jordan et al., 2019). This package allows efficient computation of the scoring rules on the basis of samples from the relevant predictive distribution. The results are presented in Figures S11–S13 for each value of $h \in \{1, 2, 4, 8\}$ and $m \in \{3, 10, 20\}$ and for every model-prior combination as well as the frequentist analysis. Each score was calculated as an average across the $40 - h + 1$ time points at which it was evaluated. The posterior predictive distributions can readily be sampled for the four Bayesian analyses. For the maximum likelihood analysis, estimation of the forecast distributions, either by simulation or numerical methods, is much more difficult. The (asymptotic) variance of the forecast errors can, in principle, be computed using a recursive algorithm that allows for epistemic uncertainty in the parameter values (see, for example Tsay, 2014, Chapter 2). However, the algorithm requires a value for the asymptotic variance of the estimator $\hat{\boldsymbol{\Phi}}$. An approximation to the variance matrix of $\{\text{vec}(\hat{A}_1), \ldots, \text{vec}(\hat{A}_p), \text{vech}(\hat{\Sigma})\}^{\text{T}}$ is available as the negative Hessian matrix evaluated at the MLE; this is a by-product of the Quasi-Newton Raphson algorithm used to minimize the negative log-likelihood function. However, converting it to a variance matrix for $\hat{\boldsymbol{\Phi}}$ requires the Jacobian of the transformation between parameter sets, analytic computation of which is prohibitively difficult. We therefore follow

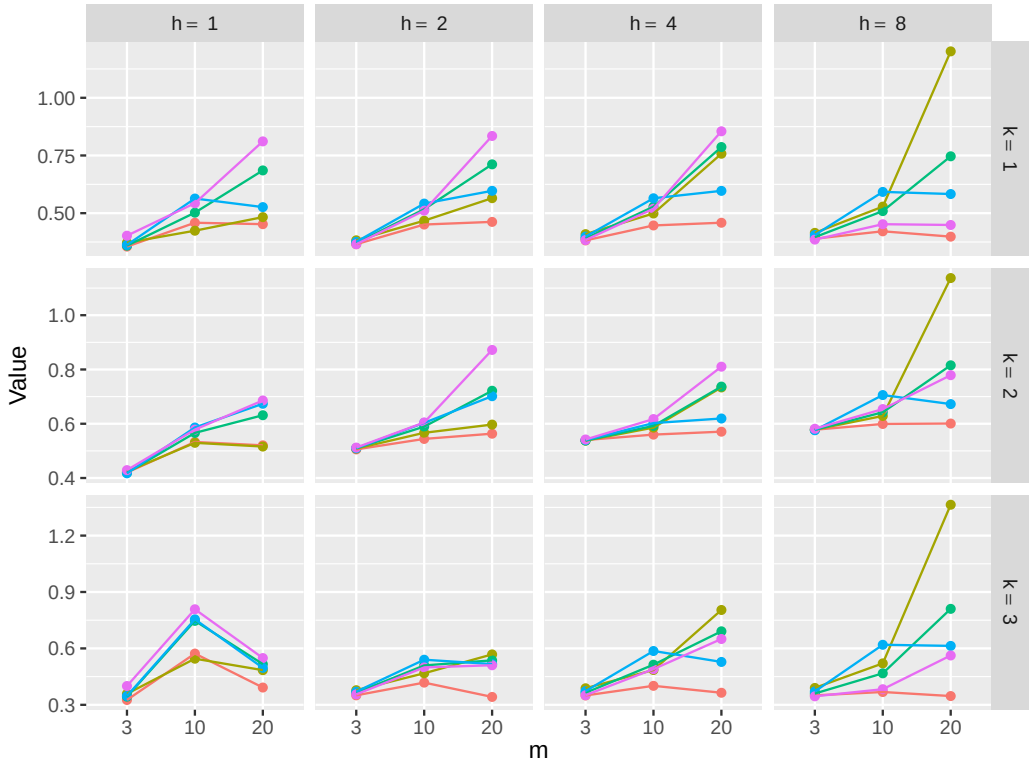


Figure S11: For each value of m, each forecast horizon h, and each prior: CRPS for each variable of interest k. The priors are: exchangeable and stationary (●); Minnesota (●); semi-conjugate (●); vague and stationary (●). Also shown are analogous scores based on forecast distributions evaluated at the stationary MLE (Ansley and Kohn, 1986) (●).

common practice in the frequentist time series literature (the so-called "plug-in" method, Pourahmadi, 2001, Chapter 2) and ignore the epistemic uncertainty when constructing h-step ahead forecast distributions. At time t, this yields a multivariate normal distribution with mean and variance equal to $E(\boldsymbol{Y}_t|\boldsymbol{y}_{1:(t-h)}, \hat{\Phi})$ and $\text{Var}(\boldsymbol{Y}_t|\boldsymbol{y}_{1:(t-h)}, \hat{\Phi}, \hat{\Sigma})$ both of which can be computed recursively (see, for example Lütkepohl, 2005, Chapter 2). The difficulty in allowing for parameter uncertainty in the forecast distributions is, of course, a substantial drawback to the maximum likelihood approach.

A discussion of the results from Figures S10–S13 can be found in Section 5 of the manuscript.

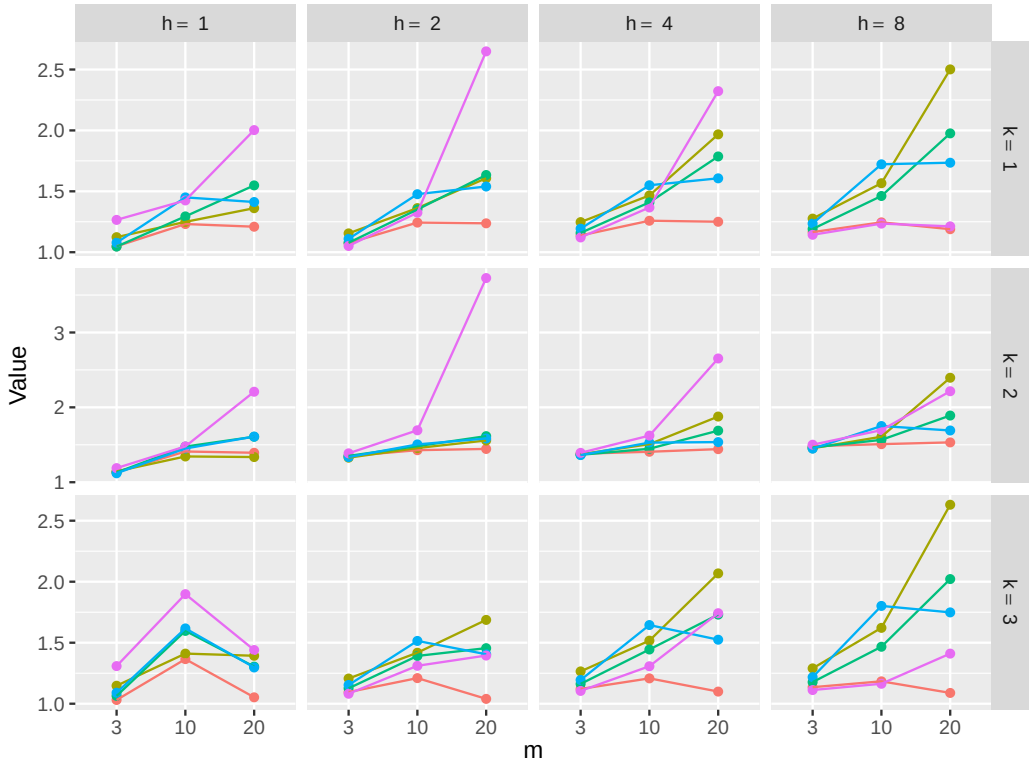


Figure S12: For each value of m, each forecast horizon h, and each prior: logarithmic score for each variable of interest k. The priors are: exchangeable and stationary (●); Minnesota (●); semi-conjugate (●); vague and stationary (●). Also shown are analogous scores based on forecast distributions evaluated at the stationary MLE (Ansley and Kohn, 1986) (●).

S7 Inference for VARMA models

As discussed in Section 6 of the manuscript, the ideas behind the symmetric partial autocorrelation parameterization and its associated prior, can be extended for VARMA models in order to constrain inference to the invertible, as well as stationary, region. Consider the VARMA$_m$(p,q) model

$$\theta(B)\boldsymbol{\epsilon}_t = \phi(B)\boldsymbol{y}_t,$$

where $\phi(B) = I_m - \phi_1 B - \cdots - \phi_p B^p$, $\theta(B) = I_m + \theta_1 B + \cdots + \theta_q B^q$ and $\boldsymbol{\epsilon}_t \sim \mathrm{N}_m(\mathbf{0}, \Sigma)$ is a white noise sequence. The stationarity condition $(\phi_1, \ldots, \phi_p) \in \mathcal{C}_{p,m}$ is handled as previously, by reparameterizing in terms of partial autocorrelation matrices $(P_1, \ldots, P_p) \in \mathcal{V}_m^p$.

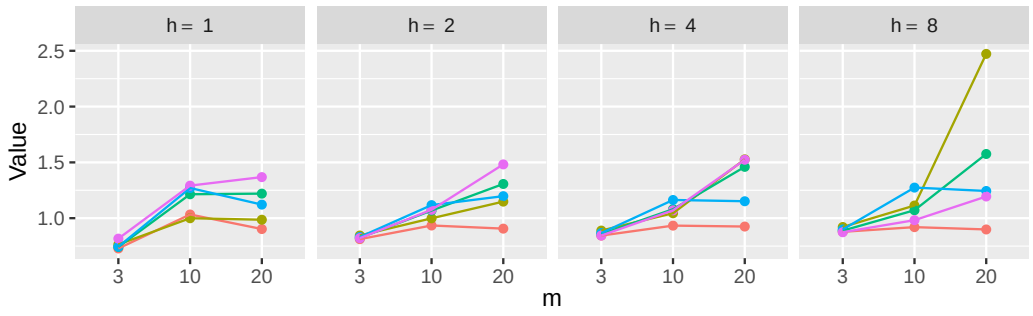


Figure S13: The energy score (ES) for each value of m, each forecast horizon h, and each prior. The priors are: exchangeable and stationary (●); Minnesota (●); semi-conjugate (●); vague and stationary (●). Also shown are analogous scores based on forecast distributions evaluated at the stationary MLE (Ansley and Kohn, 1986) (●).

Similarly, the invertibility condition $(\theta_1, \ldots, \theta_q) \in \mathcal{C}_{q,m}$ is handled by reparameterizing in terms of an analogous set of matrices $(R_1, \ldots, R_q) \in \mathcal{V}_m^q$. The recursions described in the Appendix of the manuscript are applied twice; once as if we had a pure $\text{VAR}_m(p)$ model with coefficients $\phi_1, \ldots, \phi_p$ and variance Σ to get $P_1, \ldots, P_p$, and again as if we had a pure $\text{VAR}_m(q)$ model with coefficients $-\theta_1, \ldots, -\theta_q$ and variance Σ to get $R_1, \ldots, R_q$. Unfortunately, compared with VAR processes, the parameters are more difficult to interpret. For example, the matrices Γ_s $(s = 0, 1, 2, \ldots)$ computed as a bi-product of the forward or reverse mappings for each set of parameters do not represent the autocovariance function of the VARMA process. More importantly, the P_s do not represent the partial autocorrelation matrices of the VARMA process; instead, they represent the partial autocorrelations of a model with the same autoregressive operator $\phi(B)$, but with $q = 0$. Similarly, the R_s represent a multivariate analogue of the inverse partial autocorrelation function (Bhansali, 1983) for a model with the same moving average operator $\theta(B)$, but with $p = 0$. In each case, the second transformation from Section 2.2 of the manuscript can be used to map the parameter set to unconstrained Euclidean space, $A_1, \ldots, A_p \in M_{m\times m}(\mathbb{R})^p$ and $D_1, \ldots, D_q \in M_{m\times m}(\mathbb{R})^q$. We can then assign a prior of the form

$$\pi(\Sigma, A_1, \ldots, A_p, D_1, \ldots, D_q) = \pi(\Sigma) \prod_{s=1}^{p} \pi\{\text{vec}(A_s^{\mathrm{T}})\} \prod_{s=1}^{q} \pi\{\text{vec}(D_s^{\mathrm{T}})\}$$

with the A_s and D_s assigned multivariate normal priors of the form discussed in Section 3 of the manuscript.

In order to calculate the likelihood for a VARMA model, it is convenient to introduce latent variables $\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{1-p}$ and $\boldsymbol{\epsilon}_0, \ldots, \boldsymbol{\epsilon}_{1-q}$ to initialize the process. The joint density of $(\boldsymbol{y}_0^{\mathrm{T}}, \ldots, \boldsymbol{y}_{1-p}^{\mathrm{T}}, \boldsymbol{\epsilon}_0^{\mathrm{T}}, \ldots, \boldsymbol{\epsilon}_{1-q}^{\mathrm{T}})^{\mathrm{T}}$ can be deduced from the representation of the $\mathrm{VARMA}_m(p,q)$ process as a $\mathrm{VAR}_{m(p+q)}(1)$ model; see, for example, Chapter 11 of Lütkepohl (2005). Under this representation, the role of the observation vector is played by $(\boldsymbol{y}_t^{\mathrm{T}}, \ldots, \boldsymbol{y}_{t-p}^{\mathrm{T}}, \boldsymbol{\epsilon}_t^{\mathrm{T}}, \ldots, \boldsymbol{\epsilon}_{t-q}^{\mathrm{T}})^{\mathrm{T}}$ and the single autoregressive coefficient matrix is a block matrix defined by

$$\tilde{\phi} = \left(\begin{array}{cccc|cccc} \phi_1 & \ldots & \phi_{p-1} & \phi_p & \theta_1 & \ldots & \theta_{q-1} & \theta_q \\ I_m & \ldots & 0_m & 0_m & 0_m & \ldots & 0_m & 0_m \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_m & \ldots & I_m & 0_m & 0_m & \ldots & 0_m & 0_m \\ \hline 0_m & \ldots & 0_m & 0_m & 0_m & \ldots & 0_m & 0_m \\ 0_m & \ldots & 0_m & 0_m & I_m & \ldots & 0_m & 0_m \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_m & \ldots & 0_m & 0_m & 0_m & \ldots & I_m & 0_m \end{array}\right).$$

Similarly, the error variance matrix is a block matrix defined by

$$\tilde{\Sigma} = \left(\begin{array}{cccc|cccc} \Sigma & 0_m & \ldots & 0_m & \Sigma & 0_m & \ldots & 0_m \\ 0_m & 0_m & \ldots & 0_m & 0_m & 0_m & \ldots & 0_m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_m & 0_m & \ldots & 0_m & 0_m & 0_m & \ldots & 0_m \\ \hline \Sigma & 0_m & \ldots & 0_m & \Sigma & 0_m & \ldots & 0_m \\ 0_m & 0_m & \ldots & 0_m & 0_m & 0_m & \ldots & 0_m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_m & 0_m & \ldots & 0_m & 0_m & 0_m & \ldots & 0_m \end{array}\right).$$

For a zero-mean VARMA process, the stationary mean of $(\boldsymbol{y}_t^{\mathrm{T}}, \ldots, \boldsymbol{y}_{t-p}^{\mathrm{T}}, \boldsymbol{\epsilon}_t^{\mathrm{T}}, \ldots, \boldsymbol{\epsilon}_{t-q}^{\mathrm{T}})^{\mathrm{T}}$ is clearly a length-$m(p+q)$ vector of zeros and the stationary variance $\tilde{\Gamma}_0$ can be calculated by solving the discrete Lyapunov equation

$$\tilde{\Gamma}_0 = \tilde{\phi}\tilde{\Gamma}_0\tilde{\phi}^{\mathrm{T}} + \tilde{\Sigma}$$

using vectorization and Kronecker product operators. Therefore we take

$$(\boldsymbol{y}_0^{\mathrm{T}}, \ldots, \boldsymbol{y}_{1-p}^{\mathrm{T}}, \boldsymbol{\epsilon}_0^{\mathrm{T}}, \ldots, \boldsymbol{\epsilon}_{1-q}^{\mathrm{T}})^{\mathrm{T}} \sim \mathrm{N}_{m(p+q)}(\mathbf{0}, \tilde{\Gamma}_0).$$

For further details, see Chapter 2 of Lütkepohl (2005).

Conditional on the model parameters, the joint conditional density of $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ given $(\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{1-p}, \boldsymbol{\epsilon}_0, \ldots, \boldsymbol{\epsilon}_{1-q})$ is then

$$p(\boldsymbol{y}_{1:n} \mid \boldsymbol{y}_{(1-p):0}, \boldsymbol{\epsilon}_{(1-q):0}) = \prod_{t=1}^{n} \phi_m(\boldsymbol{y}_t \mid \boldsymbol{\mu}_t, \Sigma),$$

where $\phi_m(\boldsymbol{y} \mid \boldsymbol{\mu}, \Sigma)$ denotes the density of a multivariate normal $\mathrm{N}_m(\boldsymbol{\mu}, \Sigma)$ distribution evaluated at $\boldsymbol{y}$, and where the $\boldsymbol{\mu}_t$ are defined in a forward recursion through

$$\boldsymbol{\mu}_t = \begin{cases} \sum_{i=1}^{p} \phi_i \boldsymbol{y}_{t-i} + \sum_{i=1}^{q} \theta_i \boldsymbol{\epsilon}_{t-i}, & \text{for } t = 1, \\ \sum_{i=1}^{p} \phi_i \boldsymbol{y}_{t-i} + \sum_{i=1}^{t-1} \theta_i (\boldsymbol{y}_{t-i} - \boldsymbol{\mu}_{t-i}) + \sum_{i=t}^{q} \theta_i \boldsymbol{\epsilon}_{t-i}, & \text{for } t = 2, \ldots, q, \\ \sum_{i=1}^{p} \phi_i \boldsymbol{y}_{t-i} + \sum_{i=1}^{q} \theta_i (\boldsymbol{y}_{t-i} - \boldsymbol{\mu}_{t-i}), & \text{for } t = q+1, \ldots, n. \end{cases}$$

A Stan program for fitting a zero-mean VARMA$_m$(p,q) model is given online in another supplementary file.

References

Ansley, C. F. and R. Kohn (1986). A note on reparameterizing a vector autoregressive moving average model to enforce stationarity. *J. Statist. Comput. Simul. 24*, 99–106.

Bhansali, R. J. (1983). The inverse partial correlation function of a time series and its applications. *J. Mult. Anal. 13*(2), 310–327.

Dawid, A. P. (1981). Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika 68*(1), 265–274.

Eaton, M. L. (1989). Group invariance applications in statistics. In *Regional Conference Series in Probability and Statistics, Volume 1*, pp. 1–133. Institute of Mathematical Statistics.

Edelman, A. and N. R. Rao (2005). Random matrix theory. *Acta Numerica 14*, 233–297.

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *Int. J. Forecasting 13*, 281–291.

Jauch, M., P. D. Hoff, and D. B. Dunson (2021). Monte Carlo simulation on the Stiefel manifold via polar expansion. *J. Comput. Graph. Stat. 30*(3), 622–631.

Jordan, A., F. Krüger, and S. Lerch (2019). Evaluating probabilistic forecasts with scoringRules. *J. Statist. Software 90*(12), 1–37.

Koop, G. and D. Korobilis (2009). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics 3*(4), 267–358.

Lindstrom, M. J. and D. M. Bates (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J. Amer. Statist. Assoc. 83*(404), 1014–1022.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag.

Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory*. John Wiley & Sons.

Roy, A., T. S. McElroy, and P. Linton (2019). Constrained estimation of causal, invertible VARMA. *Statist. Sinica 29*, 455–478.

Tsay, R. S. (2014). *Multivariate Time Series Analysis with R and Financial Applications*. John Wiley & Sons.